

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:29:30 ; Search time 60.99 Seconds  
(without alignments)  
119.849 Million cell updates/sec

Title: US-09-485-951-2  
1917  
Perfect score:  
Sequence: 1 MAFSGSQAPYLSPAVPFSGT.....LPTINRLEVGGDIQLTHVQT 355

Scoring table: BIOSUM62  
Gapext 0.5  
19739 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:  
1: /cgn2\_6/pctodata/1/1aa/5A\_COMB.pep:  
2: /cgn2\_6/pctodata/1/1aa/5B\_COMB.pep:  
3: /cgn2\_6/pctodata/1/1aa/6A\_COMB.pep:  
4: /cgn2\_6/pctodata/1/1aa/6B\_COMB.pep:  
5: /cgn2\_6/pctodata/1/1aa/PCTUS\_COMB.pep:  
6: /cgn2\_6/pctodata/1/1aa/backfile1.pep:  
[

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	1633	85.2	311	3	US-08-946-914-4		Sequence 4, Appli
2	783	40.8	149	2	US-08-788-584-3		Sequence 3, Appli
3	661	34.5	145	2	US-08-788-584-1		Sequence 1, Appli
4	603.5	31.5	324	3	US-08-946-914-11		Sequence 11, Appli
5	574	29.9	323	1	US-08-469-667-16		Sequence 16, Appli
6	574	29.9	323	3	US-08-046-914-2		Sequence 2, Appli
7	574	29.9	323	3	PCT-US5-07289-16		Sequence 16, Appli
8	535	27.9	145	2	US-08-788-584-5		Sequence 5, Appli
9	535	27.9	145	3	US-08-946-914-12		Sequence 12, Appli
10	472.5	24.6	316	4	US-09-131-648-5		Sequence 5, Appli
11	471.5	24.6	317	3	US-08-046-914-6		Sequence 6, Appli
12	454.5	23.7	316	2	US-08-728-521-3		Sequence 3, Appli
13	454.5	23.7	316	3	US-08-647-960-2		Sequence 2, Appli
14	454.5	23.7	316	3	US-08-646-914-15		Sequence 15, Appli
15	454.5	23.7	316	3	US-08-646-914-17		Sequence 17, Appli
16	353.5	18.4	264	2	US-08-728-521-1		Sequence 1, Appli
17	328.5	17.1	264	1	US-08-046-914-4		Sequence 4, Appli
18	321.5	16.8	336	2	US-09-131-648-1		Sequence 1, Appli
19	321	16.7	262	3	US-08-946-914-14		Sequence 6, Appli
20	318.5	16.6	250	3	US-08-046-914-10		Sequence 10, Appli
21	317.5	16.6	250	1	US-08-046-914-12		Sequence 2, Appli
22	252.5	13.3	200	3	US-08-046-914-8		Sequence 8, Appli
23	254.5	13.3	177	2	US-08-047-960-6		Sequence 6, Appli
24	250	13.0	147	2	US-08-047-960-7		Sequence 7, Appli
25	230.5	12.0	136	3	US-08-946-914-13		Sequence 13, Appli
26	226.5	11.8	135	2	US-08-047-960-5		Sequence 5, Appli
27	198.5	10.4	132	2	US-08-047-960-11		Sequence 11, Appli

#### ALIGNMENTS

RESULT 1  
US-08-946-914-4  
; Sequence 4, Application US/08946914  
; Patent No. 6027916  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Genitz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steine, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946, 914  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028, 093  
; FILING DATE: 09-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36, 688  
; REFERENCE/DOCKET NUMBER: 1488-0560001/EKS/SGW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-946-914-4  
; Query Match Score 85.2%; Score 1633; DB 3; Length 311;  
; Best Local Similarity 87.6%; Pred. No. 3.2e-164;  
; Matches 311; Conserv 0; Mismatches 0; Indels 44; Gaps  
; Qy 1 MAFSGSQAPYLSPAVPFSGTQGGLQDGLQTGVNGTVLSSSGTRFAVNFQTGFSGNDIAF 60



Qy 339 INRLEYGGDIQLTHVQT 355  
 Db 129 INRLEYGGDIQLTHVQT 145

RESULT 4  
 US-08-946-914-11  
 ; Sequence 11, Application US/08946914  
 ; Patent No. 6027916  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ni, Jian  
 ; APPLICANT: Genz, Reiner L.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Galedictin 8, 9, 10 and 10SV  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
 ; STREET: 1100 New York Ave., Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3934  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/946,914  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/028,093  
 ; FILING DATE: 09-OCT-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Steffe, Eric K.  
 ; REGISTRATION NUMBER: 36,688  
 ; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-2600  
 ; TELEFAX: 202-371-2540  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 324 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-946-914-11

Db 215 YVLPPTAKNLINERVKGGTGDIAFHMRIGD-CVVRNSYMGNSWGSSEKTPYN-PFGAG 272  
 Qy 304 QSFSVWILCEAHCLKVAVDGQLFEEYHRLNMLPNTNLVEGGDIOLTTHVQ 354  
 Db 273 QFFDLSSRCGTDREKVYFANGQHLEDFHREQAFQRYDMLTEKDTLSVYQ 323

RESULT 5  
 US-08-469-667-16  
 ; Sequence 16, Application US/08469667  
 ; Patent No. 573748  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Rosen, Craig  
 ; TITLE OF INVENTION: Colon Specific Genes and Proteins  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
 ; ADDRESS: Stewart & Oistain  
 ; STREET: 6 Becker Farm Road  
 ; CITY: Roseland  
 ; STATE: NJ USA  
 ; ZIP: 07068-1739  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/469,667  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ferraro, Gregory D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 325800-435  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 323 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-469-667-16

Query Match 29.9%; Score 574; DB 1; Length 323;  
 Best Local Similarity 31.5%; Score 603.5; DB 3; Length 324;  
 Matches 139; Conservative 54; Mismatches 115; Indels 43; Gaps 10;

Qy 9 PYLSPAVPFSGTQIGGLQDGLOITVNGTVLSSGTREAVNFTGFS-GNDIAFHMRIGD 67  
 Db 11 PTYNPTLPYKRPPIGGLVSGLSIYIOG-IAKDNMRREHVNPAVGQDEGADAFHENPRED 69

Qy 68 DGYYVVCNTROGSWGEERKTHMPFQKGMPFDLCLFLVQOSSDFKVNNGLFEVOYFHRVP 127  
 Db 70 GWDKVVENTLQDGKWWGEERKRSMPFKGAAFLFVYFLAEBYKVVVNGNFYEVCHRLP 129

Qy 70 GWDKVVENTM0QGWGKREKKSMPFKGHHFELVAMWSRYKVVVNGTFEYEGHRLP 129  
 Db 128 FHRVDLTSVNGEVQLSYTIFQNPRTVYVQPAESTVPSQVCFPPRGRRQKPPGVWPA 187  
 Qy 128 FHRVDLTSVNGSVQLSYISFQNPRTVYVQPAESTVPSQVCFPPRGRRQKPPGVWPA 187  
 Db 130 LQMYTHLQDGDLELGSINF----LGQOPAQSO1PGTM1-----PA 167

Qy 188 NPAPITQTVIHTVQSAFGMSTPAI--PPMMPHDPAYMPEFTTLLGGLYPSKSILLSG 245  
 Db 166 YPDPG----GHCPO----QLNLSLPTMGPQGPNTNP----PVPYFGRLQGLTARRTIIGK 212

Qy 168 YP-----SAGYNNPQMMSLPVNAGPPFNPNP----PVPYVGLTQGGLTARRTIIKG 214  
 Db 213 YVPTGKSFALNEVKGGSGDIALHINPRMGNGSTVVRNLLNGSGSEKKITHN - PFGPG 271

Qy 246 TVPLSAQRFHIN-LCSGNHTAFLHNRDFENAVRNTQIDNSWGSSEERSLPRKMPFVRG 303

QY 304 QSPSWILCEAHLCKVADGQHILFEYYHRLRNLPINTNLEGGDIQLTHQ 354  
 Db 272 QFFDSLIRCLDRLFKVYANGQHLDFAHRLSAFQRVDLTLIGDVTLSYVQ 322

RESULT 6  
 US-08-946-914-2  
 ; Sequence 2, Application US/08946914

; GENERAL INFORMATION:  
 ; APPLICANT: Ni, Jian  
 ; APPLICANT: Gentz, Reiner L.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Galectin 8, 9, 10 and 10Sv  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Steiner, Kessler, Goldstein, & Fox P.L.L.C.  
 ; STREET: 1100 New York Ave., Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3934  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/946,914  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/0228,093  
 ; FILING DATE: 09-OCT-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Steife, ERIC K.  
 ; REGISTRATION NUMBER: 36,688  
 ; REFERENCE/DOCKET NUMBER: 1488-0560001/BKS/SSW  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-2540  
 ; TELEFAX: 202-371-2540  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 323 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-946-914-2

Db 213 YYPPTGKSFAINFKVKGSSGDIALHINPRMGNGTUVRNSSLNGSWGSEEEKITHN-PFGPG 271  
 QY 304 QFSVWILCEAHLCKVADGQHILFEYYHRLRNLPINTNLEGGDIQLTHQ 354  
 Db 272 QFFDSLIRCLDRLFKVYANGQHLDFAHRLSAFQRVDLTLIGDVTLSYVQ 322

RESULT 7  
 PCT-US95-07289-16  
 ; Sequence 16, Application PC/TUSS9507289  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yu, Guo-Liang  
 ; APPLICANT: Rosen, Craig  
 ; TITLE OF INVENTION: Cc1on specific Genes and Proteins  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,  
 ; STREET: 6 Becker Farm Road  
 ; CITY: Roseland  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 07066-1739  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US95/07289  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ferraro, Gregory D.  
 ; REGISTRATION NUMBER: 36,134  
 ; REFERENCE/DOCKET NUMBER: 325800-265  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-994-1700  
 ; TELEFAX: 201-994-1744  
 ; INFORMATION FOR SEQ ID NO: 16:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 323 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US95-07289-16

Query Match 29.9%; Score 574; DB 5; Length 323;  
 Best Local Similarity 38.2%; Pred. No. 1.9e-52;  
 Matches 134; Conservative 52; Mismatches 121; Indels 44; Gaps 10;

QY 9 PYLSPAVPFSGTQGGLQDGQLTIVNGTUVLSSGTRPAVNFTGFS-GNDIAFHENPREF 67  
 Db 11 PTNPTLPPYQPTPGGLNVGMSVYIQQ-VASHMKRKEFVNFVGQDGSDFVAFHENPREF 69

QY 68 DGGYVVCNTRQNSWGPBPERKTHMPFQKGMPDLCFLVQSSDFKVMVNGILEVQYHFRVP 127  
 Db 70 GNIKVVFTNLOGKKGSEERKSMPFKGAAELVFTVLAEYKVVVNGNPYEVYHRLP 129

QY 128 FHRVDTISVNGSVQLSYTSFQNPRTVVPQAFSTVPSQVCPFRGRQKPPGYWPA 187  
 Db 130 LQMVTHLQVDGLQDQLQSTNFGQ-----PLRPGQ---PPMAPP 165

QY 188 NPAPITQTVIHTYQASQGMFSTPAI--PPMMYMPHAPYPMPPITTLGGLYPSKSILLSG 245  
 Db 166 YPSP-----GRHQ----QLNSLPLTMEGPPTNP---PPVPGRLQGGTARTRILKG 212

QY 246 TVLPSAQRFHIN-LCSGNHIAFHNLNPRDENAVRNTQIDNSWGSBERSLPRKMPFVRG 303  
 Db 213 YYPPTGKSFAINFKVKGSSGDIALHINPRMGNGTUVRNSSLNGSWGSEEEKITHN-PFGPG 271

QY 304 QFSVWILCEAHLCKVADGQHILFEYYHRLRNLPINTNLEGGDIQLTHQ 354

RESULT 8

Db 272 QFDLSTRCGLDRFKVYANGHQLFDFAHLRLSAFQRVDTLEIQLGDTVLTSVQ 322

Sequence 5, Application US/08788584

Patent No. 5837493

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

ATTORNEY/AGENT INFORMATION:

APPLICANT: Goli, Surya K.

APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.

APPLICANT: Petthory, Joanne R.

TITLE OF INVENTION: NOVEL HUMAN GALECTINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/788, 584

FILING DATE: Filed Herewith

CLASSIFICATION: 436

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0192 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 145 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLOSE: 727176

US-08-788-584-5

Query Match 208 FSTPAIPPMWPHPAVMPMPFTITLGLGLPKSILLSGTVPSAORHINLCSGNHIAFH 267

Best Local Similarity 70.38; Pred. No. 7.2e-19; Mismatches 9; Indels 6; Gaps 1;

Matches 104; Conservative 9; MisMatches 29; Indels 6; Gaps 1;

Db 58 LNPRFDENAVRNTQINNSWGPPEERLPSGMPFSRGQRFSTVILCGVVLSDAKRFQINURCGGDAFH 57

Query Match 268 LNPRFDENAVRNTQIDNSWGPSEERSLPRKMFVVRQSFVSYWILCEAHCLKVAVDQQLHF 327

Best Local Similarity 70.38; Pred. No. 7.2e-19; Mismatches 9; Indels 6; Gaps 1;

Matches 104; Conservative 9; MisMatches 29; Indels 6; Gaps 1;

Db 118 EYSHRLNLPDINTLEVAGDQQLTHYET 145

Query Match 328 EYVHRLRNLPPTINRLEVGGDQQLTHYET 355

Best Local Similarity 70.38; Pred. No. 7.2e-19; Mismatches 9; Indels 6; Gaps 1;

Matches 104; Conservative 9; MisMatches 29; Indels 6; Gaps 1;

Db 118 EYSHRLNLPDINTLEVAGDQQLTHYET 145

RESULT 9

US-09-131-648-5

Sequence 5, Application US/09131648

Patent No. 616920

GENERAL INFORMATION:

APPLICANT: Hallman, Jennifer L.

ATTORNEY/AGENT INFORMATION:

APPLICANT: Yee, Henry

APPLICANT: Coley, Neil C.

APPLICANT: Gieger, Karl J.

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS

FILE REFERENCE: PF-0576 US

CURRENT APPLICATION NUMBER: US/09/131,648  
 CURRENT FILING DATE: 1998-08-10  
 NUMBER OF SEQ ID NOs: 5  
 SOFTWARE: PERL Program  
 SEQ ID NO 5  
 LENGTH: 316  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE: -  
 OTHER INFORMATION: 91932712  
 US-09-131-648-5

Query Match 24.6%; Score 472.5; DB 4; Length 316;  
 Best Local Similarity 33.0%; Pred. No. 9.2e-42; Indels 51; Gaps 8;  
 Matches 115; Conservative 55; Mismatches 128; Indels 51; Gaps 8;

Qy 12 SPAVPESGNTQGGGLQDGQIQTGVNTGTVLSSGTRFAVNQFTGSGN--DIAFHNPREFD 68  
 Db 13 NPVLPFVGTTIPDQDGPGLIVRGHV-PSDAFRQVDLQNGSSVKPRADVATHFNPRFKR 71

Qy 69 GGYVVCNTRONGSWGPEERKTHMPQFGMPEFLCFLYQSSDRKVMNGILFVQYFHRYPF 128  
 Db 72 AGCIVCNLNEKWRGEEELTYDTPKREKSFELIVMLDKDQAVANGKHTLGYHRIGP 131

Qy 129 HRVDTISYNGSVQLSLSYISFQNQPRTPVQPAFSTVPPFSOPVCPPRGRQKPGW PAN 188  
 Db 132 EKDTLGIYGVNHTSGF-----SFSSLDLQST-----QAS 162

Qy 189 PAPITQTYTHYQSGQMFSTPAIIPMMYHPPAYPMPFITLGGGLYPSKSTLLSGTVL 248  
 Db 163 SLELTEIVRNPKS----GTPQL-----SLPFAARLNTPMGPGRIVVVQGEVN 207

Qy 249 PSAGRFHNLCSG--NHTAFHNPREFDENAIVRNTQDTSNSQSEERSIUPRKBFVRQSF 306  
 Db 208 ANAKSFNDLACKSKDIALNPRLNKAFTRNSEILOESNEEERNI-TSPFFSPGMYF 266

Qy 307 SWILCEAHCLKVAVDGHOLFEYHRLNLPNTIRLEGGDQLTHYQT 355  
 Db 267 EMLIYCDVREFKYAVNGVHSLEYKHFREKLSSIDTLEINGDTHLLEVRs 315

RESULT 11

US-08-946-914-6  
 ; Sequence 6, Application US/08946914  
 ; Patent No. 6027916  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ni, Jian  
 ; APPLICANT: Gentz, Reiner L.  
 ; APPLICANT: Ruben, Steven M.  
 ; TITLE OF INVENTION: Galectin 8, 9, 10 and 10sv  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Steiner, Kessler, Goldstein, & Fox P.L.L.C.  
 ; STREET: 1100 New York Ave., Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3934  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/946,914  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/028,093  
 ; FILING DATE: 09-OCT-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Steffe, Eric K.

Query Match 24.6%; Score 471.5; DB 3; Length 317;  
 Best Local Similarity 33.0%; Pred. No. 1.2e-41; Mismatches 128; Indels 51; Gaps 8;  
 Matches 115; Conservative 55; Mismatches 128; Indels 51; Gaps 8;

Qy 12 SPAVPESGNTQGGGLQDGQIQTGVNTGTVLSSGTRFAVNQFTGSGN--DIAFHNPREFD 68  
 Db 14 NPVLPFVGTTIPDQDGPGLIVRGHV-PSDAFRQVDLQNGSSVKPRADVATHFNPRFKR 72

Qy 69 GGYVVCNTRONGSWGPEERKTHMPQFGMPEFLCFLYQSSDRKVMNGILFVQYFHRYPF 128  
 Db 73 AGCIVCNLNEKWRGEEELTYDTPKREKSFELIVMLDKDQAVANGKHTLGYHRIGP 132

Qy 129 HRVDTISYNGSVQLSLSYISFQNQPRTPVQPAFSTVPPFSOPVCPPRGRQKPGW PAN 188  
 Qy 130 HRVDTISYNGSVQLSLSYISFQNQPRTPVQPAFSTVPPFSOPVCPPRGRQKPGW PAN 188  
 Db 133 EKDTLGIYGVNHTSGF-----SFSSLDLQST-----QAS 163

Qy 189 PAPITQTYTHYQSGQMFSTPAIIPMMYHPPAYPMPFITLGGGLYPSKSTLLSGTVL 248  
 Db 164 SLELTEIVRNPKS----GTPQL-----SLPFAARLNTPMGPGRIVVVQGEVN 208

Qy 249 PSAGRFHNLCSG--NHTAFHNPREFDENAIVRNTQDTSNSQSEERSIUPRKBFVRQSF 306  
 Db 209 ANAKSFNDLACKSKDIALNPRLNKAFTRNSEILOESNEEERNI-TSPFFSPGMYF 267

Qy 307 SWILCEAHCLKVAVDGHOLFEYHRLNLPNTIRLEGGDQLTHYQT 355  
 Db 268 EMLIYCDVREFKYAVNGVHSLEYKHFREKLSSIDTLEINGDTHLLEVRs 316

RESULT 12  
 US-08-748-521-3  
 ; Sequence 3, Application US/08728521  
 ; Patent No. 5869389  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; APPLICANT: Hawkins, Phillip R.  
 ; TITLE OF INVENTION: HUMAN GALECTIN HOMOLOG  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Drive  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ Version 1.5  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/728,521  
 ; FILING DATE: Filed Herewith  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0137 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 SEQUENCE FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 316 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 IMMEDIATE SOURCE:  
 LIBRARY: Genbank  
 CLONE: 71107032

US-08-728-521-3

Query Match 23.7%; Score 454.5; DB 2; Length 316;  
 Best Local Similarity 31.7%; Pred. No. 7.3e-10;  
 Matches 114; Conservative 62; Mismatches 133; Gaps 8;

Qy 1 MAFSGSQAPYLSPAYPFSGTQGLQDGLQTIVNGTLYLSSGTRAVNFGTS---GND 57  
 Db 2 LSLSNLQNLINNPTIPVYSTITEQLKPGSLIVRSHVPKDS-ERQVDFOHGSNLKPRAD 60  
 Qy 58 IAFHENPREFDGGYVVCNTRONSGMPPEERKTHMPQKGMFPDLFCLVQSSDFEKUWNGI 117  
 Qy 61 WAFHENPREFRSNCIVCNLTNEKGWSWELTHDMFRKEKSFEIVMVLKRNKFHVAVNGK 120  
 Db 118 LFVQYFHRVPFHRDVTISVNGSYLISFQNPRTFVQPAFSTVPSQFCFPPRGR 177  
 Qy 121 HILLYAHRINPEKIDTLGIFGKVNHSIGFR-----FSSDLQSM----- 159  
 Db 178 RQKPPGWWPANPAPITQVIAHQSAGQMFSTPAIPMMYPHPAYPMPPITITLGGLYP 237  
 Qy 160 -----ETSTLGLTQISKENIQS-GKLHL-----LPPEARLNASMGP 196  
 Qy 238 SKSILLSGTVLPSSAQRFHINLCSG--NHIAPHLNPRDENAVVRNTQIDNSWGSEERSLP 295  
 Db 197 GRTVYVKGEYNTNAFSFNDLVAQGSRDIAHHLNPLNVAFVRNSFLQDAGGEERNI- 255  
 Qy 296 RKMPFVRQGQFSVWILCEAHCLKVAVDQHLEFEYVHRLNLPINTNLEVGDDIOLTHVQT 355  
 Db 256 TCFPFSSGMYFEMIYCDVREFKVAVNGVHSLEYKRFKDLSSTLAVGDDIRLDVRS 315

RESULT 13

US-08-647-960-2  
 Sequence 2, Application US/08647960  
 Patent No. 5908761

## GENERAL INFORMATION:

APPLICANT: ZICK, Yehiel  
 TITLE INVENTION: GALECTIN-8 AND GALECTIN-8-LIKE PROTEINS  
 NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BROWDY AND NEIMARK  
 STREET: 419 Seventh Street, N.W., Suite 300  
 CITY: Washington  
 STATE: D.C.  
 ZIP: 20004

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/647,960  
 FILING DATE: 30-MAY-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: IL 107880  
 FILING DATE: 05-DEC-1993

ATTORNEY/AGENT INFORMATION:  
 NAME: BROWDY, Roger L.  
 REGISTRATION NUMBER: 25,618  
 REFERENCE/DOCKET NUMBER: ZICK-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-628-5197  
 TELEFAX: 202-737-3527  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 316 amino acids  
 MOLECULE TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-647-960-2

Query Match 23.7%; Score 454.5; DB 2; Length 316;  
 Best Local Similarity 31.7%; Pred. No. 7.3e-40;  
 Matches 114; Conservative 62; Mismatches 133; Gaps 8;

Qy 1 MAFSGSQAPYLSPAYPFSGTQGLQDGLQTIVNGTLYLSSGTRAVNFGTS---GND 57  
 Db 2 LSLSNLQNLINNPTIPVYSTITEQLKPGSLIVRSHVPKDS-ERQVDFOHGSNLKPRAD 60  
 Qy 58 IAFHENPREFDGGYVVCNTRONSGMPPEERKTHMPQKGMFPDLFCLVQSSDFEKUWNGI 117  
 Db 61 WAFHENPREFRSNCIVCNLTNEKGWSWELTHDMFRKEKSFEIVMVLKRNKFHVAVNGK 120  
 Qy 118 LFVQYFHRVPFHRDVTISVNGSYLISFQNPRTFVQPAFSTVPSQFCFPPRGR 177  
 Db 121 HILLYAHRINPEKIDTLGIFGKVNHSIGFR-----FSSDLQSM----- 159  
 Db 178 RQKPPGWWPANPAPITQVIAHQSAGQMFSTPAIPMMYPHPAYPMPPITITLGGLYP 237  
 Db 160 -----ETSTLGLTQISKENIQS-GKLHL-----LPPEARLNASMGP 196  
 Qy 238 SKSILLSGTVLPSSAQRFHINLCSG--NHIAPHLNPRDENAVVRNTQIDNSWGSEERSLP 295  
 Db 197 GRTVYVKGEYNTNAFSFNDLVAQGSRDIAHHLNPLNVAFVRNSFLQDAGGEERNI- 255  
 Qy 296 RKMPFVRQGQFSVWILCEAHCLKVAVDQHLEFEYVHRLNLPINTNLEVGDDIOLTHVQT 355  
 Db 256 TCFPFSSGMYFEMIYCDVREFKVAVNGVHSLEYKRFKDLSSTLAVGDDIRLDVRS 315

RESULT 14

US-08-946-914-15  
 Sequence 15, Application US/08946914  
 Patent No. 6027916  
 GENERAL INFORMATION:  
 APPLICANT: Ni, Jian  
 APPLICANT: Gentz, Reiner L.  
 APPLICANT: Ruben, Steven M.  
 TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
 STREET: 1100 New York Ave., Suite 600  
 CITY: Washington  
 STATE: D.C.  
 ZIP: 20005-3034  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/946,914  
 FILING DATE: Herewith  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/028,093

PRIOR APPLICATION DATA:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Steffe, Eric K.  
 REGISTRATION NUMBER: 36, 688  
 REFERENCE/DOCKET NUMBER: 1488-0560001/EKS/SGW  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-371-2540  
 TELEFAX: 202-371-2540  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 316 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-946-914-15

Query Match 23.7%; Score 454.5; DB 3; Length 316;  
 Best Local Similarity 31.7%; Pred. No. 7.3e-40; Gaps 8;  
 Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps

1	MAFGSQSQA PVLSPAVP FSGSPV LQGGLQ DGLQIT VNGTVL SSGTRFA VNGTGF S--GND	57	Query Match 23.7%; Score 454.5; DB 3; Length 316; Best Local Similarity 31.7%; Pred. No. 7.3e-40; Gaps 8; Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps
Db	2 LSLSNLQNTI YNTPIPV STQITE LQKPGS T1LIVGR HVPKDS -ERQVDF QHGNSL KPRAD	60	1 MAFGSQASQ APVLSPAV PFSGSPV LQGGLQ DGLQIT VNGTVL SSGTRFA VNGTGF S--GND
Db	58 IAFHENP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117	57	2 LSLSNLQNTI YNTPIPV STQITE LQKPGS T1LIVGR HVPKDS -ERQVDF QHGNSL KPRAD
Qy	61 VAFHENP RFRKSNC LVCNTL NEKNGWE ETHDME FRKESF EIVIMV LKNKFHV AVNGK	120	3 LAFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	118 LEVQYF HRYPFH RVTDTIS NGSYQ LSYISTON PRTV PQVCF PPRPRGR	177	4 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Qy	121 HILLYAH RNPKEK IDTGLG KVNHSIGR -FSSD LQSM	159	5 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	178 RQKPPGV WPANP APITQT VHTV QSAPG QHNF STPA IAPP MMYP PHAP MPF ITL GVP	237	6 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Qy	160 E- STLGL QTQ ISKEN IQS -GK HLS	196	7 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	238 SKSILL GTVPS AQRH FHLNCS -NHTA FHNLN PREDEN AVRNT QIDNS WGS SEERS LSP	295	8 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Qy	197 GRTVV KGE VNT NAT SFN VLD VAG ERS RDL HNL PRA LYKA FVRS NSFL QOW GPE ERN -255	315	9 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	296 RKMPE VRGQSF SFSV NLI CEAH CLKY AVD GQH LF EY YH RNL MPT INR LEV GGD IQ LTH YQT	355	10 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Qy	256 TCFP SSGM YF EMI YCD VREF YAV NGV SLEY KHR FKL SSID TLA VGD DIR DVR 315	355	11 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	197 GRTVV KGE VNT NAT SFN VLD VAG ERS RDL HNL PRA LYKA FVRS NSFL QOW GPE ERN -255	315	12 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117
Db	256 TCFP SSGM YF EMI YCD VREF YAV NGV SLEY KHR FKL SSID TLA VGD DIR DVR 315	315	13 1AFHFNP RPRFEDGG YVVCNT RQNGSW GPEEKTH MPQKGP FDLGFY QOSDFK YVNGT1 117

RESULT 15  
US-08-946-914-17  
Sequence 17, Application US/08946914  
Patent No. 6027916  
GENERAL INFORMATION:  
APPLICANT: Ni, Jian  
; APPLICANT: Reiner L.  
; APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steine, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946, 914  
; FILING DATE: Herewith  
; CLASSIFICATION: 530

Scoring table:	BLOSUM62	Query:	MAFSGSQAPYLSPAVPFSGT . . . . . LPTINRLEVGGDILQTHVQT	Match:	355	Length:	12	DB:	Swissprot_39 *	ID:	
<b>Scanned:</b> 93435 seqs, 34255486 residues											
Total number of hits satisfying chosen parameters:	93435										
Minimum DB seq length:	0										
Maximum DB seq length:	2000000000										
Post-processing: Minimum Match	0%										
	Maximum Match	100%									
Listing first 45 summaries											
Database :											
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.											
<b>SUMMARIES</b>											
Result	No.	Score	Query	Match	Length	DB	ID	Description			
1	1917	100.0	355	1	LEGG_HUMAN			00182 homo sapien			
2	1374.5	71.7	354	1	LEGG_RAT			P97840 rattus norvegicus			
3	1324	69.1	353	1	LEGG_MOUSE			P08573 mus musculus			
4	603.5	31.5	324	1	LEGG_RAT			P38552 rattus norvegicus			
5	574	29.9	323	1	LEGG_HUMAN			P56470 homo sapien			
6	541	28.2	323	1	LEGG_PIG			P02058 sus scrofa			
7	535	27.9	144	1	LEGG_RAT			P47967 rattus norvegicus			
8	518	27.0	301	1	LEGG_MOUSE			P04891 mus musculus			
9	467.5	24.4	316	1	LEGG_MOUSE			Q91115 mus musculus			
10	463.5	24.2	316	1	LEGG_HUMAN			P00214 homo sapien			
11	454.5	23.7	316	1	LEGG_RAT			Q62665 rattus norvegicus			
12	427	22.3	283	1	LEGL_HAECHO			P44126 haemophilus ducreyi			
13	414.5	21.6	279	1	LE32_CAEEL			P36573 caenorhabditis elegans			
14	374	19.5	285	1	LE33_CAEEL			P019581 caenorhabditis elegans			
15	339	17.7	244	1	LEG3_CRL20			P47953 cricetulus leucurus			
16	330.5	17.2	295	1	LEG3_CANFA			P38486 canis familiaris			
17	324.5	16.9	241	1	LEG3_RABBIT			P47845 oryctolagus cuniculus			
18	322.5	16.8	263	1	LEG3_MOUSE			P16110 mus musculus			
19	321	16.7	261	1	LEG3_RAT			P02699 rattus norvegicus			
20	318.5	16.6	249	1	LEG3_HUMAN			P17931 homo sapiens			
21	230.5	12.0	135	1	LEG7_HUMAN			P47929 homo sapiens			
22	223.5	11.7	135	1	LEG7_MOUSE			P044974 mus musculus			
23	216.5	11.3	135	1	LEG7_RAT			P97590 rattus norvegicus			
24	189.5	9.9	134	1	LEG6_CHICK			P23668 gallus gallus			
25	184.5	9.6	134	1	LEG4_CHICK			P07583 gallus gallus			
26	182	9.5	182	1	LEC7_CAEEL			P09605 caenorhabditis elegans			
27	167	8.7	134	1	LEGL_MOUSE			P16045 mus musculus			
28	166	8.7	134	1	LEG1_SHEEP			P81184 ovis aries			
29	164	8.6	134	1	LEG1_CRK1R			P48538 cricetulus leucurus			
30	164	8.6	134	1	LEGL_RAT			P11762 rattus norvegicus			
31	164	8.6	184	1	LEC8_CAEEL			P09610 caenorhabditis elegans			
32	160	8.3	134	1	LEG1_BOVIN			P11116 bos taurus			
33	155	8.1	129	1	LEG_EELBEL			P08520 electrophorus electricus			

## ALIGNMENTS

RESULT	1	LEGG_HUMAN	STANDARD	PRT;	355 AA.
ID	LEGG_HUMAN				
AC	00182; O14532; O75028;				
DT	01-NOV-1997 (Rel. 35; Created)				
DT	15-JUL-1998 (Rel. 36; Last sequence update)				
DT	01-OCT-2000 (Rel. 40; Last annotation update)				
DE	GALECTIN-9 (HOM-HD-21) (ECALECTIN).				
GN	LGALS9.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX					
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT FORM).				
RC	SEQUENCE FROM Spleen;				
RA	RUEBERI O.; Schmitt H.; Padle N.; Pfreundschuh M.; Sahin U.;				
RT	"Molecular definition of a novel human galectin which is immunogenic in patients with Hodgkin's disease.";				
RL	J. Biol. Chem. 272:6416-6422(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A. (LONG FORM).				
RC	TISSUE=Gastric carcinoma;				
RA	Kato S.;				
RL	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A. (SHORT FORM).				
RX	MEDLINE-#97197815; PubMed-9045665;				
RA	Matsumoto R.; Matsumoto H.; Seki M.; Hata M.; Asano Y.; Kanegasaki S.;				
RA	Tuereci O.; Hirashima M.;				
RT	Stevens R.L.; Hirashima M.;				
RT	"Human ecallectin, a variant of human galectin-9, is a novel eosinophil chemoattractant produced by T lymphocytes.";				
RL	J. Biol. Chem. 273:16976-16984(1998).				
CC	-1- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.				
CC	CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	CC -1- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.				
CC	CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.				
CC	CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.				
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).				
CC	CC DR EMBL; Z49107; CAA88972; 1;				
CC	CC DR EMBL; AB006782; BAA2166; 1;				
CC	CC DR EMBL; AB005894; BAA31542; 1;				

DR	HSSP; P17931; 1A3K. MM; 601879; -.	Query Match	100.0%	Score 1.917; DB 1.; Length 355;
DR	InterPro; IPR001079; -.	Test Local Similarity	100.0%	Pred. No. 1..3e-146; 0; Gaps 0;
DR	PROSITE; PS00337; Gal-bind_lectin; 2.	Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
KW	Galaptin; Lectin; Repeat; Alternative splicing			
FT DOMAIN	1 148	148	LINKER.	
FT DOMAIN	149 206	355	GALAPTIN 1.	
FT BINDING	82 88	88	BETA-GALACTOSIDE (BY SIMILARITY).	
FT BINDING	287 293	MISSING (IN SHORT ISOFORM).		
FT VARSPLIC	149 180	5	G -> S (IN REF. 3).	
FT CONFLICT	88 88	88	K -> R (IN REF. 1).	
FT CONFLICT	135 135	135	S -> F (IN REF. 1).	
FT CONFLICT	270 313	270	P -> L (IN REF. 1).	
FT CONFLICT	313 313	313	E -> G (IN REF. 1).	
SEQUENCE	355 AA; 39518 MW;	4748C22FCFA536A CRC64;		
QY	1 MAFSGSQAPYLSPAVPFSGTIGQLDQGLQTIVNGTIVLSSSGTREAVNFOTGSGNDIAF 60	Query Match	100.0%	Score 1.917; DB 1.; Length 355;
Db	1 MAFSGSQAPYLSPAVPFSGTIGQLDQGLQTIVNGTIVLSSSGTREAVNFOTGSGNDIAF 60	Test Local Similarity	100.0%	Pred. No. 1..3e-146; 0; Gaps 0;
QY	61 HFNPREFDGGYVCNTRONGSGPEERKTHMFEQKGMPEDLQFLVQSSDFKVMVNGILFV 120	Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Db	61 HFNPREFDGGYVCNTRONGSGPEERKTHMFEQKGMPEDLQFLVQSSDFKVMVNGILFV 120			
QY	121 QYHRVPYHRVDTISVNGSVQQLSYISQNPRTVPVQPAFSTVPSQPVCFPPRPRRQK 180			
Db	121 QYHRVPYHRVDTISVNGSVQQLSYISQNPRTVPVQPAFSTVPSQPVCFPPRPRRQK 180			
QY	181 PPGVWPANPAPITQTVLHQYQAPGMESTPAIAPPMMYPHAYPMFITTILGGLYPSKS 240			
Db	181 PPGVWPANPAPITQTVLHQYQAPGMESTPAIAPPMMYPHAYPMFITTILGGLYPSKS 240			
QY	241 ILLSGTVLPSAQRFHINCGSNHIAFHLPNPREDENAVRNTQIDNSWGSEERSLPRKMPF 300			
Db	241 ILLSGTVLPSAQRFHINCGSNHIAFHLPNPREDENAVRNTQIDNSWGSEERSLPRKMPF 300			
QY	301 VRQSQFSEWVILCEAHCLKVAVNGQHLEFEYVHLRLNLPNTINLEVGDIQLTHVQT 355			
Db	301 VRQSQFSEWVILCEAHCLKVAVNGQHLEFEYVHLRLNLPNTINLEVGDIQLTHVQT 355			
GULT 2				
9_RAT	STANDARD; PRT; 354 AA.	Query Match	71.7%	Score 1.374;5; DB 1.; Length 354;
AC	P97840; 008588; Q35866.	Best Local Similarity	71.8%	Pred. No. 3.7e-103; Mismatches 68; Indels 1; Gaps 1;
DR	01-NOV-1997 (Rel. 35, Created)	Matches 255; Conservative 31; Mismatches 68; Indels 1; Gaps 1;		
DR	15-JUL-1998 (Rel. 36, Last sequence update)			
DR	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	GALECTIN-9 (36 kDa BETA-GALACTOSIDE BINDING LECTIN) (URATE TRANSPORTER/CHANNEL) (UAT).			
GN	LGALS9			
OS	Rattus norvegicus (Rat).			
OC	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Muriidae; Murinae; Rattus.			
OX	NCBI_TaxID-10116			
RN	[1]	SEQUENCE FROM N.A. (LONG AND SHORT FORMS).		
RC	SEQUENCE FROM N.A. (LONG AND SHORT FORMS).			
RX	STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney, and Small intestine;			
RA	Wada, J.; Kanwar, Y.S.;			
RA	"Identification and characterization of galectin-9, a novel beta-galactoside-binding mammalian lectin.";			
RL	J. Biol. Chem. 272:6078-6086 (1997).			
RN	[2]	SEQUENCE FROM N.A. (SHORT FORM).		
RP	4748C22FCFA536A CRC64;			

301 VRGOSFSWILCEAHCLKVAVDGQHLEFYHRLRNLPTRNLEVGDDIQLTTHVQT 355  
Y FT VARSPLJC 148 178 MISSING (IN SHORT ISOFORM).  
Y SQ SEQUENCE 353 AA; 40036 MW; B5403FF6E280C531 CRC64;

FT	VARSPLIC	148	178	MISSING (IN SHORT ISOFORM).
SQ	SEQUENCE	353 AA;	40036 MW;	B54036F6E280C531 CRC64;

Query Match	69.1%	Score 1324;	DB 1;	Length 353;	
Best Local Matches	69.3%	Pred. No. 4.2e-99;			
Local Matches	246;	Conservative	38;	Gaps 2;	
Matches	246;	Mismatches	69;		
QY	1	MAFGSSOAPYSPAYPFSQGTTQGGGLDQGQIYTNGTNTVLSGGTRFAVNQFGSGNDIAF	60		
Db	1	MALFSSQSPYINPIPTGTQGGQEGLOVLTQGT-T-KSFAQREFVNFQNSFNGNDIAF	59		
QY	61	HFNPREDFGDDGYYVCNTRONSGWPPEERKTHMPKGMPEFLCCLVQSSDFKVAYNGILFV	120		
Db	60	HFNPREDFGDDGYYVCNTRONSGWPPEERKTHMPKGMPEFLCCLVQSSDFKVAYNGILFV	119		
QY	121	QYHRRPVPHRTDTISNGSVOLSYLSFQNPTPTVQPAFFSTVPSQSPYCEPPPRPRGRKOK	180		
Db	120	QYHRRPVPHLDTAVSGCLKLSEITQN-SAAPPQHVSTLQSQVQFPRTKGRKOK	178		
QY	181	PPGVWPNAPAPITOTYTHYOSAPGOMESTPAIAPPMMYPHPAYPMPEFITTIGGLYPSKS	240		
Db	179	TONFRPRAHQAPMAQTTIHMWVSTPGMFSPTGIPPVYPPAYTIPFYTPIPGLYPSKS	238		
QY	241	ILLSGTVLPSAQRFHINLCSGNHIAFHHLNPREDENAVRNTQIDNSWGSERSLSPRKMPF	300		
Db	239	TMISGQSFPLDRAFTHNLRCGDTIAFHHLNPREDENAVRNTQIDNSWGSERSLSPRKMPF	298		
QY	301	VRQQSFPSWVILCEAHCLVKAVDQGQHFEYVYHRLRNPFTPIRLEVYGGDIDQIYTHVQT	355		
Db	299	SRQQSFPSWVILCEGHCEFKVAVNGQHMEYHRLKNLQDINTLEVADQIYTHVQT	353		
RESULT	4				
LIE64_RAT		STANDARD;	PRT;	324 AA.	
AC	P38552;				
DT	01-OCT-1994	(Rel. 30, Created)			
DT	01-OCT-1994	(Rel. 30, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	GALETIN-4 (LACTOSE BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)				
GN	L36LBP.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
NCBI_TAXID=	10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Intestine;				
RC	MEIDLINE=93194902; PubMed=8449956;				
RA	Oda Y., Herrmann J., Gitt M., Turck C.W., Burlingame A.L.,				
RA	Barondes S.H., Leffler H.,				
RT	"Soluble lactose-binding lectin from rat intestine with two different carbohydrate-binding domains in the same peptide chain.";				
RT	carbohydrate-binding domains				
RL	J. Biol. Chem. 268:5939-5939 (1993).				
CC	-1- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF SUGARS.				
CC	-1- SUBUNIT: MONOMER.				
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FULL-LENGTH FORM IN SMALL AND LARGE INTESTINE AND STOMACH BUT WAS NOT DETECTED IN OTHER TISSUES INCLUDING LUNG, LIVER, KIDNEY, AND SPLEEN.				
CC	-1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.				
CC	-1- SIMILARITY: BELONGS TO THE GALACTIN (S-LECTIN) FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to license@isb-sib.ch).				

Page 4

RC	TISSUE-Tongue;	QY	353 VQ 354		
RX	MEDLINE-95081129; PubMed-7989350;	DB	321 VQ 322		
RA	Chiu M.L., Parry D.A.D., Feldman S.R., Klapper D.G., O'Keefe E.J.;				
RT	"An adherens junction protein is a member of the family of lactose-				
RT	binding lectins";				
RL	J. Biol. Chem. 269:31770-31776 (1994).				
RN	[2]				
RP	SEQUENCE OF 1-140 FROM N.A.	RESULT 7			
RC	TISSUE-Small intestine;	LEGS5_RAT	STANDARD; PRT; 144 AA.		
RC	MEDLINE-96327507; PubMed-8672129;	ID	LEGS5_RAT		
RA	Wintero A.K., Fredholm M., Davies W.;	AC	P47967;		
RT	*Evaluation and characterization of a porcine small intestine cDNA	DT	01-FEB-1996 (Rel. 33, Created)		
RT	library: analysis of 839 clones.;	DT	01-FEB-1996 (Rel. 33, Last sequence update)		
RL	Mamm. Genome 7:509-517 (1996).	DT	01-FEB-1996 (Rel. 33, Last annotation update)		
CC	-1- FUNCTION: GALACTIN THAT BINDS LACTOSE AND A RELATED RANGE OF	DE	GALACTIN-5 (RL-18).		
CC	SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.	GN	LGALS5.		
CC	-1- SUBUNIT: MONOMER.	OS	Rattus norvegicus (Rat).		
CC	-1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	DOMAINS.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
CC	-1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.	NCBI_TaxID	10116;		
CC	-----	RN	[1]		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	RC	TISSUE=Reticulocytes;		
CC	the European Bioinformatics Institute. There are no restrictions on its	RX	MEDLINE-95197487; PubMed-78906611;		
CC	use by non-profit institutions as long as its content is in no way	RA	Gitt M.A., Wiser M.P., Jeffller H., Herrmann J., Xia Y.-R.,		
CC	modified and this statement is not removed. Usage by and for commercial	RA	Massa S.M., Cooper D.N.W., Luisi A.J., Barondes S.H.;		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	RT	*Sequence and mapping of galectin-5, a beta-galactoside-binding		
CC	-----	RT	lectin, found in rat erythrocytes.		
CC	-----	RL	J. Biol. Chem. 270:5032-5038 (1995).		
CC	-----	CC	-1- FUNCTION: MAY FUNCTION IN ERYTHROCYTE DIFFERENTIATION.		
CC	-----	CC	-1- SUBUNIT: MONOMER.		
CC	-----	CC	-1- TISSUE SPECIFICITY: ERYTHROCYTES.		
CC	-----	CC	-1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.		
CC	-----	CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	CC	-----		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	CC	-----		
CC	the European Bioinformatics Institute. There are no restrictions on its	CC	-----		
CC	use by non-profit institutions as long as its content is in no way	CC	-----		
CC	modified and this statement is not removed. Usage by and for commercial	CC	-----		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	CC	-----		
DR	PFam: PF00337: Gal-bind_lectin; 2.	CC	-----		
DR	PROSITE: PS00309; GALAPTIN; FALSE_NEG.	DR	EMBL; L36862; AAC42050.1; -.		
RW	Galaptin; Lectin; Repeat.	DR	EMBL; L21711; AAA65445.1; -.		
FT	DOMAIN 1 152	DR	HSSP: P17931; 1A3K.		
FT	DOMAIN 153 177	DR	InterPro; IPRO01079;		
FT	DOMAIN 178 323	DR	Pfam: PF00337; Gal-bind_lectin; 1.		
FT	BINDING 256 262	DR	PROSITE: PS00309; GALAPTIN; 1.		
FT	CONFLICT 98 99	KW	Galaptin; Lectin; Acetylation.		
FT	CONFLICT 126 126	FT	INIT_MET 0		
FT	H -> T (IN REF. 2).	FT	MOD_RES 1 1		
SEQUENCE	323 AA; 35852 MW; 7280761712B23343 CRC64;	FT	BINDING 76 82		
SEQUENCE	-----	FT	CONFLICT 127 135		
SEQUENCE	-----	SQ	MISSING (IN AAA65445).		
Query Match	Score 541; DB 1; Length 323;	SEQUENCE	144 AA; 16065 MW; BC95283D760DAS15 CRC64;		
Best Local Similarity	28.2%	Query Match	27.9%		
Matches 132; Conservative 54; Mismatches 128; Indels 48; Gaps 11;	Best Local Similarity	36.5%	Score 535; DB 1; Length 144;		
Db	1 MAF---SGSOAPYPLSPAVPFGSTGQIQLQGDLQTIVNTVLSGSGTRFAVNFQTFGFS-GN 56	Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;	Best Local Similarity	70.3%	Score 535; DB 1; Length 144;
QY	1 MAFVPAQGQPTY-NPTLPYKPLPGGLRVMGMSVYIQQ-VANEHMKRFVNVYVGQGPEA 58	Db	3 FSTQT-----PYPNLAVPFTSIPNLGYPKS1IVSGVWLSDAKRFQINLRCGGDIAPH 56	Db	3 FSTQT-----PYPNLAVPFTSIPNLGYPKS1IVSGVWLSDAKRFQINLRCGGDIAPH 56
Db	57 DIAFHENPREDGGYVVCNTRQNSGMPWGPPEERKTHMPFQKGMPFDICFLYQSISDEFKVMYNG 116	QY	268 LNPRDENEAVVRNTQIDNSMGESEERSLPRKMPFVRGQSFSWILCEAHC1KVAVDGOHLF 327	QY	268 LNPRDENEAVVRNTQIDNSMGESEERSLPRKMPFVRGQSFSWILCEAHC1KVAVDGOHLF 327
QY	59 DVAFHENPREDGWDKVFNVSQDGKWNQEKKRSMFPRKAFAFLVIMLPHEHYKVVNG 118	Db	57 LNPRDENEAVVRNTQIDNSMGESEERSLPRKMPFVRGQSFSWILCEAHC1KVAVDGOHLF 327	QY	328 EYHRHLRNLPTINRLVEGGD1QLTHVOT 355
QY	117 ILFYQYFHRYPFHRDITSYNGSYLQSISFQNRTVYQPAFSTVPEFQVPCPPRPRG 176	Db	57 EYSHRLMLNLPDINTVEAGD1QLTHVOT 144	Db	117 EYSHRLMLNLPDINTVEAGD1QLTHVOT 144
Db	119 DPFYEGFHRIPLPVQLVTHLQDGTJLQSINF----IGQGAPS-----157	QY	208 FSTPAIPMMYPHPAYPMPPFITLIGGLYPSKS1ILSGTLPYLSAQRFHINLCSGNHIAFH 267	QY	208 FSTPAIPMMYPHPAYPMPPFITLIGGLYPSKS1ILSGTLPYLSAQRFHINLCSGNHIAFH 267
QY	177 RQRKPPGVWPAANPAPITQVTVTQVOSAPGOMFSTPAIPPMYMPHAY - PMPFITLLOG 234	Db	3 FSTQT-----PYPNLAVPFTSIPNLGYPKS1IVSGVWLSDAKRFQINLRCGGDIAPH 56	Db	3 FSTQT-----PYPNLAVPFTSIPNLGYPKS1IVSGVWLSDAKRFQINLRCGGDIAPH 56
Db	158 -----PGMP-NPG ----- YPGPKHNNQPCNLFQMEGAPTFNPVPPYKTRLQGG 201	QY	235 LYPSKS1LLGTVPLPSAQRFHIN--LCSGNHIAFPHLNPRDENEAVVRNTQIDNSMGESEER 292	QY	235 LYPSKS1LLGTVPLPSAQRFHIN--LCSGNHIAFPHLNPRDENEAVVRNTQIDNSMGESEER 292
QY	202 LVARRTIVKGYVPPSGKSLVINFKVQSSDVALHINPRTEGIVRNSYLNKGWAEER 261	Db	293 SLPRMMPFVRGQSFYWIICAEH1KVAVDGQHFEYXYLRLNPLTINLEVGSDIQLQH 352	QY	293 SLPRMMPFVRGQSFYWIICAEH1KVAVDGQHFEYXYLRLNPLTINLEVGSDIQLQH 352
Db	262 KSSFN-PFAFGQYFQDLSIRGDRFKVYANGQHLDQFVDTLEIQGdvtLSY 320	LEG6_MOUSE	RESULT 8	LEG6_MOUSE	

ID	LEGG6_MOUSE; STANDARD;	PRT;	301 AA.	Qy	188	NPAPITOTVHTVQASAPGOMESTPAIPIPMYYPHAYPMPMFITTLGGLYPSKSILLSGTV	247
AC	088352;			Db	160	-----AMTG-----	1
DT	15-DEC-1998 (Rel. 37, Created)			Qy	248	LPQAQFHIN-LC5SNHIAFHNPREFDNVQFTRKMPFVQGOS	305
DT	15-DEC-1998 (Rel. 37, Last sequence update)			Db	194	LPTAKTFAINFRVGSSEDTALHINPRIG CLVRISYMMGSWGTEFRMAYN-PFGPGQF	251
DE	GALECTIN-6.			Qy	306	FSWVILCEAHLCKVADQHLEFEYHHLRNLNPTINRLEYGGDIOLTHY	353
GN	LGALS6			Db	252	FDLISRGMDRKEVFKFANGIHLFENFSHREQLRKINTLENGDLTLSYV	299
OS	Mus musculus (Mouse);						
OC	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;						
OX	NCBI_TaxID=10090;						
RN	[1]						
RP	SEQUENCE FROM N. A.						
RC	RESTRN-129/5V;						
RX	MEDLINE=98112847; PubMed=9446608;						
RA	Gitt M.A., Colnot C., Poirier F., Nani K.J., Barondes S.H.,						
RA	Leffler H.,						
RT	"Galectin-4 and galectin-6 are two closely related lectins expressed in mouse gastrointestinal tract."						
RT	in mouse gastrointestinal tract."						
RL	J. Biol. Chem. 273:2954-2960(1998).						
RA	Gitt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,						
RA	Leffler H.,						
RT	"Sequence, structure, and chromosomal mapping of the mouse Lgals6 gene, encoding galectin-6."						
RN	273:2954-2960(1998).						
RL	Chem. 273:2954-2960(1998).						
CC	-1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS						
CC	-1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).						
CC	EMBL; AF026799; AAC04508.1; -.						
DR	EMBL; AF026796; AAC04508.1; JOINED.						
DR	EMBL; AF026797; AAC04508.1; JOINED.						
DR	EMBL; AF026798; AAC04508.1; JOINED.						
DR	EMBL; AF026794; AAC27244.1; -.						
DR	MGI; MGI:107535; Igals6.						
DR	InterPro; IPR001079; -.						
DR	Pfam; PF00337; Gal-bind_lectin; 2.						
DR	Galaptin; Lectin; Repeat.						
DOMAIN	1	151	GALAPTIN 1.				
DOMAIN	152	160	LINKER.				
FT	160	301	GALAPTIN 2.				
FT	154	154	A -> V (IN AAC27244).				
SEQUENCE	301 AA;	34112 MW;	9A4DD09944EDFAB9 CRC64;				
SQ				Score 518; DB 1; Length 301;			
Query Match	27.0%			Score 467.5; DB 1; Length 316;			
Best Local Similarity	35.6%			Best Local Similarity 32.4%; Pred. No. 1.6e-30;			
Matches	124; Conservative			Matches 113; Conservative 61; Mismatches 124; Indels 51; Gaps 8;			
Qy	9	PYLSPAVPFSGTIIQGGLQDGQLQITVNGTVLSSGTRAVNFQTG-FSGNDIAFHNPREF	67	Qy	12	SPAVPFSGTIIQGGLQDGQLQITVNGTVLSSGTRAVNFQTGFS--GNIDIAFHNPREF	68
Db	11	PTYNPTLPYKRIIPGGLSVGNMSFYIQLTA-KENMRFFHVNAFGQDDGAVFHNPREFD	69	Db	13	NPLIPIVGTITEQLKGSLIVTRGHVPKDS-ERFQDFQLNSLXPRADYAFHNPREFR	71
Qy	68	DGGYVWNTRQNGSWGPEERKTHMPFQKGMPFDLCLFVQSSDFKIVWNGILFVQYFHRVP	127	Qy	69	GGVVVNTRQNGSWGPEERKTHMPFQKGMPFDLCLFVQSSDFKIVWNGILFVQYFHRVP	128
Db	70	GidkVvFNTKQSGRMGKEEKS-MPFQKGKHFELYFWMVMPHEKHYKVVNGSPFYEGHRLP	128	Db	72	SSCVVNTLQEKGWEETYDMPFFKEKSEIVPVALNKRFQAVNGRIVLYAHRISP	131
Qy	128	FHRVDTISVNGSVQLSYISFQNPRTPVQPAFSTVPSQVCPFPFRGRQKPPGQWPA	187	Qy	129	HRVDTISVNGSVQLSYISFQNPRTPVQPAFSTVPSQVCPFPFRGRQKPPGQWPA	188
Db	129	WOMVTHLQVGDUELOSINF----FGVQPAETKYP-----	159	Db	132	EQIDTVGIGYGRNVIHSIGFR-----FSSDLQSM-----	162



RP	SEQUENCE FROM N.A.	DE	32 KDa BETA-GALACTOSIDE-BINDING LECTIN (GALECTIN 1).
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;	GN	GAL-1.
RX	MEDLINE=95155445; PubMed=7852431;	OS	Haemonchus contortus.
RA	Hadari Y.R., Paz K., Dekel R., Mestrovic T., Accili D., Zick Y.;	OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
RT	"Galactin-8. A new rat lectin, related to galectin-4."	OC	Trichostomylida; Haemonchidae; Haemonchus.
RL	J. Biol. Chem. 270:3447-3453(1995).	NCBIL_TaxID=6289;	OX
CC	-!- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.	[1]	RN
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.		RN
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, CARDIAC MUSCLE, LUNG, AND BRAIN.		RP
CC	-!- DEVELOPMENTAL STAGE: VERY LOW LEVELS IN WHOLE EMBRYOS, HIGH LEVELS IN ADULT TISSUES.		RP
CC	-!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.		RP
CC	-!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.		RP
CC	This SWISS-PROT entry is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).		RP
CC	DR EMBL: AAA66359; 1; -.	DR	EMBL: AF036098; AAB88823; 1; -.
DR	DR InterPro: IPR01079; -.	DR	InterPro: IPR01079; -.
DR	DR Pfam: PF00337; Gal-bind_lectin; 2.	DR	Pfam: PF00337; Gal-bind_lectin; 2.
DR	DR PROSITE: PS00309; GALAPTIN; 2.	DR	PROSITE: PS00309; GALAPTIN; 2.
KW	KW Galaptin; Lectin; Repeat.	KW	Galaptin; Lectin; Repeat.
FT	DOMAIN 1 153 GALAPTIN 1.	FT DOMAIN	1 150 GALAPTIN 1.
FT	DOMAIN 154 184 LINKER.	FT DOMAIN	151 283 GALAPTIN 2.
FT	DOMAIN 185 316 GALAPTIN 2.	FT BINDING	217 223 BETA-GALACTOSIDE (BY SIMILARITY).
FT	BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).	FT	SEQUENCE 283 AA; 32530 MW; 84D19BFF383750 CRC64;
SQ	SEQUENCE 316 AA; 36038 MW; C04B766CCE913D59 CRC64;	SQ	
Query Match	23.7%; Score 454.5; DB 1; Length 316;	Query Match	22.3%; Score 427; DB 1; Length 283;
Best Local Similarity	31.7%; Pred. No. 1.8e-29; Mismatches 133; Indels 51; Gaps 8;	Best Local Similarity	30.4%; Pred. No. 2.5e-27; Mismatches 107; Indels 82; Gaps 10;
Matches	114; Conservative 62; Mismatches 133; Indels 51; Gaps 8;	Matches	106; Conservative 54; Mismatches 107; Indels 82; Gaps 10;
Qy	1 MAFSGSQAPYLSPAYFEGIYVVCNTRGNSGPEERKTHMFQKGMPFEDLCFLQVSSGTFAVNFOTGFS --GND 57	Qy	10 YLSPAVPFESTIIOGGIQLQITVNGTIVLSSGTGFAVNQT -- GFSQGNDIAFHFNPREF 66
Db	2 LSLSNQNLIYNTIYPTIYVSTITEQLPKGSILIVRHVPKDS-ERQVDFQHGNSLKPRAD 60	Db	11 YNKP-VYPSLQLQEITPGLVKSTIDS-QRTFTNLHSKSAFSDGNDVPLHSIVRF 68
Qy	58 IAFHENPREFENGYYVVCNTRGNSGPEERKTHMFQKGMPFEDLCFLQVSSGTFAVNFOTGFS --GND 57	Qy	67 EDGGIVVVCNTRGNSGPEERKTHMFQKGMPFEDLCFLQVSSGTFAVNFOTGFS --GND 57
Db	61 VAFHENPREFKNSNCIVTNTKNEKGWEETHDMPFRKESFETIVMVKNKEHIVAVNGK 120	Db	69 -DEGVKVMNTFANGEMKGKEERKS-LPIKIGDSFDTRIRADDRFIVTIVDKEFDRYHRL 126
Qy	118 LFVQYFHRVPPHRVDTISVNGSVQLSYISQNPRIVPVQPAFSTYVPSQVCFPPRPRGR 177	Qy	127 PFHRYDTISYNGSYVOLSYISFQNPRIVPVQPAFSTYVPSQVCFPPRGRQQPKGVNP 186
Db	121 HILLYAHRINPEKIDTLGIFSKVNLHSIGR-FSSDLQSM----- 159	Db	127 PLTSITHLSIDGDLNHNHVN----- 148
Qy	178 RQKPPGVWPAAPATITQTVIHTVQSSAPQMFESTPAIIPPMYPHPAYPMPPITITLGLYP 237	Qy	187 ANPAPITQTIVHTVQSAQGMESTPAIIPPMYPHPAYPMPPITITLGLYPSKSTLSGT 246
Db	160 -----ETSTLGLTQISKENIQKS-GKLHS-----LFPEARLNASMGP 196	Db	149 GK----- 149
Qy	238 SISLISLGTQVLPSSAOFPHINICSG-NHIAFHNPREFDENAVVRQTSQNGWSEERSL 295	Qy	176 VEKKAKRFENNLRRNGDIALHENPREFDEAVIRNALAANEWGNEERE - GKMPEFKVG 233
Db	197 GRTVVVKGEVNTNATSFNDVAGRSRSDIALHNPRLNPKVAFVRSNFSLQDAGPEERNI - 255	Qy	306 FSVWILCEAHLCKVAVDQHLEFYHRLRNULPTINRLVEGGDQIQLTHVQ 354
Qy	296 RKMPPFVRGQSFSWVILCEAHCLKVAVDQHLEFYHRLRNULPTINRLVEGGDQIQLTHVQ 355	Db	234 FDLAIKNEAYAFQIFVNGERETSFAHR-QDPNDISGLQIQGDIELTGIQ 281
Db	256 TCFPSSGMYFEMITYCDVREFKVAVNGVHSLEYKHRFKDLSIDTLAVGDTRLLDVR 315	Db	
RESULT	12	RESULT	13
LE3.2_CAEEL	LE3.2_CAEEL	LE3.2_CAEEL	LE3.2_CAEEL
ID	ID	ID	ID
LEG1_HAEAO	STANDARD;	STANDARD;	STANDARD;
ID	LEG1_HAEAO	LEG1_HAEAO	LEG1_HAEAO
AC	AC	AC	AC
04126;	P3657;	P3657;	P3657;
DT	01-JUN-1994	01-JUN-1994	01-JUN-1994
30-MAY-2000 (Rel. 39, Created)	(Rel. 29, Created)	(Rel. 29, Last sequence update)	(Rel. 29, Last sequence update)
DT	01-OCT-2000 (Rel. 40, Last annotation update)	01-OCT-2000 (Rel. 40, Last annotation update)	01-OCT-2000 (Rel. 40, Last annotation update)
DE	3.2 KDa BETA-GALACTOSIDE-BINDING LECTIN (32 kDa GBP).	3.2 KDa BETA-GALACTOSIDE-BINDING LECTIN (32 kDa GBP).	3.2 KDa BETA-GALACTOSIDE-BINDING LECTIN (32 kDa GBP).
GN	W09H1.6.	W09H1.6.	W09H1.6.



67 IYNAVTKTGKEERAKN-PIKKGDDDFDTRIRAHDSKFQVNSINKEVKNFEHRIPLN SV 125  
 Db 132 DTISVGNSVOLSYISQNPRVPPVQPAFSTVPSQVCFPRRRQKPGVNPANPAP 191  
 Qy 126 SHLSIDGQDVVLNHYQ ----- -WGGK-- 144  
 Db 192 ITQTVIHTVQSGQMESTPAIPPMYPHPAYPMPPITLIG-GLYPSKSILLSTVLP S 250  
 Qy 145 ----- YPPVYESGIAADGLVPGKTLVWGTPEKK 174  
 Db 251 AQRFHINLCSGN-HIAFHHLNPRFDENAVVRNTQDNWSGSEERSLPLRKMPFVRGOSFSY W 309  
 Qy 175 AKFNFNLLKNGNDIALHFNPRDFEKSVRNSLVNGEWNRE--GKNDFERLTAFDLE 232  
 Db 310 IICEAICLKVAVDQGQHLEIYHRMLRMLPTINRLEVGGDIOLTHVQ 354  
 Qy 233 IRNEEFAFQIFVNGERFASTAHV-DPHDIAGLQIQGDIELTGQI 276

RESULT 15

CRIL0 STANDARD; PRT; 244 AA.

AC P47953; 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE GALECTIN-3 (GALACTOSE-SPECIFIC LECTIN 3) (MAC-2 ANTIGEN) (IGE-BINDING PROTEIN) (35 kDa LECTIN) (CARBOHYDRATE BINDING PROTEIN 35) (CBP 35)  
 DE (LAMININ-BINDING PROTEIN) (LECTIN L-29) (CBP30).  
 GN LGALS3  
 OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster);  
 Mammalia; Eutheria; Rodentia; Craniata; Vertebrata; Euteleostomi;  
 OC Cricetidae; OC  
 OC Cricetus; OC  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE=94399546; PubMed=8027086;  
 RA Mehl B., Bawumia S., Martin S.R., Hughes R.C.;  
 RT "structure of baby hamster kidney carbohydrate-binding protein CBP30, an S-type animal lectin";  
 RL J. Biol. Chem. 269:18250-18258(1994).  
 CC -!- FUNCTION: GALACTOSE-SPECIFIC LECTIN WHICH BINDS IGE.  
 CC -!- SIMILARITY: THE C-TERMINAL DOMAIN BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; X78879; CAA55479.1; - .  
 DR HSSP; P17931; IAK.  
 DR InterPro; IPR01079; - .  
 DR Pfam; PF00337; Gal-bind\_lectin; 1.  
 DR PROSITE; PS00309; GALAPTIN; 1.  
 FT DOMAIN 34 98 Galaptin; Lectin; IGE-binding protein; Repeat; phosphorylation; Acetylation.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT MOD\_RES 5 5 PHOSPHORYLATION (BY CK1)  
 FT PROSITE; PS00309; GALAPTIN; 1.  
 FT DOMAIN 34 98 7 X 9 AA, TANDEM REPEATS OF Y-P-G-X(3)-P-  
 FT REPEAT 34 42 1.  
 FT REPEAT 43 51 2.  
 FT REPEAT 52 60 3.  
 FT REPEAT 61 69 4.

Scoring table:	BLOSUM62				Alignments			
Scorers:	Gapop 10.0 , Gapext 0.5				RESULT 1			
Searched:	93435 seqs, 34255486 residues				LEGG9_HUMAN	STANDARD;	PRT;	355 AA.
Total number of hits satisfying chosen parameters:	93435				ID LEG9_HUMAN	075038;		
Minimum DB seq length: 0	AC 0018B; O14532; O19012; A2AB_ERIEU				AC 0018B;	O14532;		
Maximum DB seq length: 20000000000	DT 01-NOV-1997 (Rel. 35, Created)				DT 01-NOV-1997	(Rel. 35,		
Post-processing: Minimum Match 0%	DT 15-JUL-1998 (Rel. 36, Last sequence update)				DT 15-JUL-1998	(Rel. 36,		
Post-processing: Maximum Match 100%	DT 01-OCT-2000 (Rel. 40, Last annotation update)				DT 01-OCT-2000	(Rel. 40,		
Database :	DE GALECTIN-9 (HOM HD-21) (ECALECTIN).				DE GALECTIN-9	(HOM HD-21)		
Sequence:	GN LGALS9.				GN LGALS9.			
Title:	Homo sapiens (Human).				OS Homo sapiens (Human).			
Perfect score:	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				OC Eukaryota; Metazoa;			
Sequence:	Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.				OC Mammalia;			
1 US-09-485-951-1	RN 90423 glycine max				OC Butheria;			
1 NVRTVPVQPAFSTVPSQVPCFPVPRGRQQK 32	RN 006649 mus musculu				OC Primates;			
	RN 094751 schizosaccharomyces pombe				OC Catarrhini;			
	RN 000423 glycine max				OC Hominidae;			
	RN 094751 schizosaccharomyces pombe				OC Homo.			
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							
	RN 094751 schizosaccharomyces pombe							
	RN 000423 glycine max							

DR HSSP; P17931; 1A3K.  
 DR MIM; 601879; -.  
 DR InterPro; IPR01079; -.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 KW Galaptin; Lectin; REPEAT; Alternative splicing.  
 DOMAIN 1 148 GALAPTIN 1.  
 FT DOMAIN 149 206 LINKER.  
 FT DOMAIN 207 355 GALAPTIN 2.  
 FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 287 293 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPLIC 149 180 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 5 5 G > S (IN REF. 3).  
 FT CONFLICT 88 88 K > R (IN REF. 1).  
 FT CONFLICT 135 135 S > F (IN REF. 1).  
 FT CONFLICT 270 270 P > L (IN REF. 1).  
 FT CONFLICT 313 313 E > G (IN REF. 1).  
 SEQUENCE 355 AA; 39518 MW; 4748C22FCAPA536A CRC64;

Query Match 100.0%; Score 178; DB 1; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-15;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0; Gaps 0; Gaps 0;

Qy 1 NPRTVPVQPAFSTVPPFSQPVCPPRPRGRQQK 32  
 Db 149 NPRTVPVQPAFSTVPPFSQPVCPPRPRGRQQK 180

RESULT 2  
 LEG9\_RAT  
 ID LEG9\_RAT STANDARD; PRT; 354 AA.  
 AC P97840; 008588; O58666;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-9 (36 kDa BETA-GALACTOSIDE BINDING LECTIN) (URATE  
 TRANSPORTER/CHANNEL) (URAT).  
 GN GALS9.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Murinae; Mus.  
 OX [1]  
 RN RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).  
 RX STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney, and Small intestine;  
 RX MEDLINE=97190351; PubMed=9038233;  
 RA Wada J.; Kanwar Y. S.;  
 RA "Identification and characterization of galectin-9, a novel beta-  
 galactoside-binding mammalian lectin.";  
 RT "Molecular cloning and functional reconstitution of a urate  
 transporter/channel.";  
 RL J. Biol. Chem. 272:617-625(1997).  
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-  
 EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY  
 SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS  
 CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE  
 METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC  
 TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND  
 GASTROINTESTINAL URATE EXCRETION HIGHLY SELECTIVE TO THE ANION  
 URATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-  
 CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPlicing.  
 CC -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPlicing.

CC SMALL INTESTINE.  
 CC -!- DOMAINS: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC -!  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -!  
 CC -!- AA51192; 1; -.  
 DR EMBL; 059462; AAB8592; 1; -.  
 DR EMBL; 067958; AAB8591; 1; -.  
 DR InterPro; IPR001079; -.  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR PROSITE; PS00309; GalAPTIN; 2.  
 KW Galaptin; Lectin; REPEAT; Alternative splicing; Ion transport.  
 FT DOMAIN 1 147 GALAPTIN 1.  
 FT DOMAIN 148 205 LINERIN.  
 FT DOMAIN 206 354 GALAPTIN 2.  
 FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPLIC 148 179 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 354 AA; 3946 MW; 6574F960B2EAFF37C CRC64;

Query Match 58.4%; Score 104; DB 1; Length 354;  
 Best Local Similarity 59.4%; Pred. No. 5.1e-06; Indels 9; Gaps 0; Gaps 0;

Matches 19; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 NPRTVPVQPAFSTVPPFSQPVCPPRPRGRQQK 32  
 Db 148 N斯塔APVQPVFSTMQFSQPVCPPRPRGRKQR 179

RESULT 3  
 LEG9\_MOUSE  
 ID LEG9\_MOUSE STANDARD; PRT; 353 AA.  
 AC 008573; 008572; 1  
 AC 008573; 008572; 2  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE GALS9.  
 GN GALS9.  
 OS Mus musculus (Mouse).  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX [2]  
 RN NCBI\_TAXID=10116;  
 RN RP SEQUENCE FROM N.A.  
 RN SEQUENCE FROM N.A. (LONG AND SHORT FORMS).  
 RN STRAIN=CD-1; TISSUE=Small intestine, and Kidney;  
 RN MEDLINE=97190351; PubMed=9038233;  
 RA Wada J.; Kanwar Y. S.;  
 RA Abramson R. G.;  
 RT "Identification and characterization of galectin-9, a novel beta-  
 galactoside-binding mammalian lectin.";  
 RL J. Biol. Chem. 272:6078-6086(1997).  
 CC -!  
 CC SEQUENCE FROM N.A. (SHORT FORM).  
 CC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;  
 CC MEDLINE=97150769; PubMed=8985305;  
 CC Leal-Pinto E., Tao W., Rapaport J., Richardson M., Knorr B. A.,  
 CC Abramson R. G.;  
 CC RT "Molecular cloning and functional reconstitution of a urate  
 transporter/channel.";  
 CC RL J. Biol. Chem. 272:617-625(1997).  
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-  
 EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY  
 SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS  
 CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE  
 METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC  
 TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND  
 GASTROINTESTINAL URATE EXCRETION HIGHLY SELECTIVE TO THE ANION  
 URATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-  
 CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPlicing.  
 CC -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPlicing.

- TISSUE SPECIFICITY: ACCENTUATED EXPRESSION IN LIVER AND THYMUS OF EMBRYO. DEFECTED IN EMBRYONIC HEART, BRAIN, LUNG, LIVER, AND KIDNEY. HIGHLY EXPRESSED IN ADULT THYMUS, SMALL INTESTINE, AND LIVER, AND TO A LESSER EXTENT IN LONG, KIDNEY, SPLEEN, CARDIAC, AND SKELETAL MUSCLE. BARELY DETECTABLE IN BRAIN AND RETICULOCYTE. THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
- DEVELOPMENTAL STAGE: THE EXPRESSION INCREASED WITH SUCCESSIVE STAGES OF EMBRYONIC DEVELOPMENT.
- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING DOMAINS.
- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annual> or send an email to license@isb-sib.ch).

EMBL: U55061; AAB51190.1; -  
 EMBL: U55060; AAB51189.1; -  
 HSSP: P17331; 1A3K.  
 MGD: MGI:094961; 9aals9.  
 InterPro: IPR001079; -  
 Pfam: PF00337; Gal-bind\_lectin; 2.  
 PROSITE: PS00309; GALAPTIN\_2.  
 Galaptin; Lectin; Repeat; Alternative splicing.

DOMAIN	1	147	GALAPTIN 1.
DOMAIN	148	204	LINKER
DOMAIN	205	353	GALAPTIN 2.
BINDING	81	87	BETA-GALACTOSIDE (BY SIMILARITY).
BINDING	285	291	BETA-GALACTOSIDE (BY SIMILARITY).
VARSPLIC	148	178	MISSING (IN SHORT ISOFORM).
SEQUENCE	353 AA;	40036 MW;	B54036F6280C531 CRC64;

Query Match 52.8%; Score 94; DB 1; Length 353;  
 Best Local Similarity 66.7%; Fred. No. 9.8e-05;  
 Matches 18; Conservative 3; Mismatches 6; Indels 0; Gaps

SEQUENCE	6 PVQPAFESTVPSQVCEPPRGRQRQK 32	111 111: 111 11 1:11:11 152 PVQHVESTLQFSQVQEPRTPKGRQRQK 178
----------	--------------------------------	---

RESULT 4

SEQUENCE	07_MOUSE	STANDARD:	PRT:	806 AA.
SEQUENCE FROM N.A.	MK07_MOUSE			
SEQUENCE S., Moriguchi T., Nishida E.;	Q9WVS8;			
Activation of the protein kinase ERK5/ERMK1 by receptor tyrosine	01-OCT-2000 (Rel. 40, Created)			
kinases: identification and characterization of a signalling pathway to	01-OCT-2000 (Rel. 40, Last sequence update)			
the nucleus."	01-OCT-2000 (Rel. 40, Last annotation update)			
MITOGEN-ACTIVATED PROTEIN KINASE 7 (EC 2.7.1. -) (EXTRACELLULAR SIGNAL-REGULATED KINASE 5) (ERK-5) (ERMK1 KINASE).	MAPK7 OR ERK5.			
Mus musculus (Mouse).				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCBI_TaxID:10090;				

[1] SEQUENCE FROM N.A.

Kamakura S., Moriguchi T., Nishida E.;

Activation of the protein kinase ERK5/ERMK1 by receptor tyrosine kinases: identification and characterization of a signalling pathway to the nucleus."

Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS (BY SIMILARITY).

-!- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE PHOSPHORYLATION (BY SIMILARITY).

-!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.



Rus musculus (Mouse).  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
NCBI\_TaxID=10090;

[1] SEQUENCE FROM N.A. (LONG FORM).  
STRAIN=C57BL/6; TISSUE=Liver;  
MEDLINE=94344118; PubMed=1065341;  
Reisz-Porszasz S., Probst M.R., Fukunaga B.N., Hankinson O.;  
"Identification of functional domains of the aryl hydrocarbon  
receptor nuclear translocator protein (ARNT)." ;  
Mol. Cell. Biol. 14:6075-6086(1994).  
[2] SEQUENCE FROM N.A. (SHORT FORM).  
STRAIN=C57BL/6;  
MEDLINE=95050586; PubMed=7961746;  
LI H., Dong L., Whitlock J.P. Jr.;  
"Transcriptional activation function of the mouse Ah receptor nuclear  
translocator." ;  
J. Biol. Chem. 269:28098-28105(1994).  
- I - FUNCTION: REQUIRED FOR ACTIVITY OF THE AH (DIOXIN) RECEPTOR. THIS  
PROTEIN IS REQUIRED FOR THE LIGAND-BINDING SUBUNIT TO TRANSLOCATE  
FROM THE CYTOSOL TO THE NUCLEUS AFTER LIGAND BINDING. THE COMPLEX  
THEN INITIATES TRANSCRIPTION OF A GENES INVOLVED IN THE ACTIVATION  
OF PAH PROCARCINOGENS.  
- I - SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
BHLH PROTEIN. FORMS AN HETEROODIMER WITH AHR, WITH HIF1A AS WELL AS  
WITH OTHER BHLH PROTEINS.  
- I - SUBCELLULAR LOCATION: NUCLEAR.  
- I - TISSUE SPECIFICITY: Ubiquitous.  
- I - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
TRANSCRIPTION FACTORS.  
- I - SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

[2] RN SEQUENCE OF 43-61 AND 76-87.  
RC TISSUE-Muscle;  
RA Kluxen F.-W., Vandekerckhove J., Schoeffl F., Jockusch H.;  
RL Mouse News Lett. 84: 76-77 (1989).  
CC -1- SUBUNIT: INTERACTS WITH C-TERMINAL DOMAIN OF ACTIN-BINDING PROTEIN  
CC 280.  
CC -1- SPECIFICITY: FOUND IN BOTH CARDIAC AND SLOW SKELETAL  
CC (SOLEUS) MUSCLE.  
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
CC FAMILY.  
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AAF155909; AAF20023.1; -.  
DR MGD; MGI:1352494; Hsp25-2.  
DR InterPro; IPR001436; -.  
DR InterPro; IPR002068; -.  
DR Pfam; PF00011; HSP20.1.  
DR PRINTS; PR00299; ACRYSTALLIN.  
DR PROSITE; PS01031; HSP20.1.  
KW Heat shock; Chaperone.  
DOMAIN 17 30 POLY-SER.  
FT CONFLICT 60 61 EP -> GD (IN REF. 2).  
SQ SEQUENCE .169 AA; 18621 MW; 120F4ED8EA9D0E7C CRC64;

Query Match 28.9%; Score 51.5%; DB 1; Length 169;  
Best Local Similarity 33.3%; Pred. No. 7.9;  
Matches 11; Conservative 6; Mismatches 7; Indels 9; Gaps 1;

QY 6 PVPQAFS-----TVFSPQVCPFPPRPRGR 29  
Db 39 PMEKALSMFSDDFGSFMPHSEPLAFFPARPGQQ 71

RESULT 10  
PSU1\_YEAST  
ID PSU1\_YEAST  
AC P53350;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
MRNA DECAPPING PROTEIN 2 (PSU1 PROTEIN).  
DCCP2 OR PSU1 OR YNL118C OR N1917.  
Saccharomyces cerevisiae (Baker's yeast).  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomyces;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
NCBI\_TaxID:4932;  
RN [1]  
RN RP SEQUENCE FROM N.A.  
RN RC STRAIN="273-10B;  
RN RA Tzagoloff A.A.;  
RN RL Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RN RP SEQUENCE FROM N.A.  
RX MEDLINE=7245296; PubMed=9090055;  
RA de Antoni A.; D'Angelico M.; Dal Pero F.; Sartorello F.; Pandolfo D.;  
RA Pallavicini A.; Lanfranchi G.; Valle G.;  
RT The DNA sequence of cosmid 14-13b from chromosome XIV of  
Saccharomyces cerevisiae reveals an unusually high number of  
RT overlapping open reading frames.;  
RL Yeast 13:261-266 (1997).  
RN [3]  
RN RP INTERACTION WITH DCP1.  
RX PubMed=10508173;

RA Durckley T.; Parker R.; required for mRNA decapping in Saccharomyces cerevisiae and contains a functional Mutt motif.";  
RT RT  
RL EMBO J. 18:5311-5422(1999).  
CC -1- FUNCTION: REQUIRED FOR THE PRODUCTION OF ACTIVE DECAPPING ENZYME, PERHAPS IN A PROCESS REQUIRING THE HYDROLYSIS OF A PYROPHOSPHATE BOND. DECAPPING IS A THE MAJOR PATHWAY OF mRNA DEGRADATION IN YEAST. IT OCCURS THROUGH DEADENYLATION, DECAPPING AND SUBSEQUENT 5' TO 3' EXONUCLEARYTIC DECRY OF THE TRANSCRIPT BODY.  
CC -1- SUBUNIT: INTERACTS WITH DCP1.  
CC -1- SIMILARITY: BELONGS TO THE NDIX HYDROLASE FAMILY. STRONG, TO S. POMBE SPAC19A8.12.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; L43065; AAA68866.1; -.  
CC DR EMBL; 269382; CAA93389.1; -.  
CC DR EMBL; 271394; CAA5998.1; -.  
CC SGD; S0005062; DCP2.  
CC Pfam; PF00931; mutt; 1.  
CC PRINTS; PR0002; MUTTDOMAIN.  
CC DR PROSITE; PS00893; NDIX; 1.  
CC KW Hydrolase.  
CC FT DOMAIN 134 155 NDIX BOX.  
CC FT DOMAIN 436 439 POLY-SER.  
CC FT CONFLICT 425 425 P -> L (IN REF. 1).  
CC SQ SEQUENCE 970 AA: 108667 MW; D53CA2C5A546FA4A CRC64;

Query Match 28.9%; Score 51.5%; DB 1; Length 970;  
Best Local Similarity 30.4%; Pred. No. 43; Mismatches 14; Indels 17; Gaps 1;

QY 1 NPRTVPVCPAFTVP-----FSQPVCFPPRPRGR 29  
Db 386 NP1PTPVDPNFNAPPNPMAFGVBNMHNLSGPVSPQFSLPPAPLPR 431

RESULT 11  
MYCN\_MARMO  
ID MYCN\_MARMO  
AC Q61976;  
STANDARD; PRT;  
AC Q61976;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE N-MYC PROTO-ONCOGENE PROTEIN (N-MYC1).  
GN MYCN OR NMYC OR NMYC1.  
OS Marmota monax (Woodchuck).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae; Marmota.  
OC NCBI\_TaxID=9095;  
OX [1]  
RP SEQUENCE FROM N.A.  
RN TISSUE=Liver  
RX MEDLINE=90370481; PubMed=23995655;  
RA Forel G.; Tiollais P.; Buendia M.-A.;  
RT "Nucleotide sequence of the woodchuck N-myc gene (WN-myc1).";  
RL Nucleic Acids Res. 18:4918-4918(1990).  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHHL PROTEIN. BINDS DNA AS AN HETERO-DIMER WITH MAX.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX LOOP HELIX (BHHL) FAMILY OF TRANSCRIPTION FACTORS. BHHL-TIP SUBFAMILY.  
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and thus statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X53673; CAA37712.1; .

DR EMBL: X53674; CAA37712.1; JOINED.

DR InterPro; IPR00109; .

DR InterPro; IPR00148; .

DR InterPro; IPR002015; .

DR Pfam; PF00010; HUH; 1.

DR PRINTS; PRO0044; LENDIPRNYC.

DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.

RW Nuclear protein, DNA-binding; Proto-oncogene; Phosphorylation.

FT DOMAIN 258 274 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 377 390 BASIC DOMAIN.

FT DOMAIN 391 430 HELLIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 429 450 LEUCINE-ZIPPER (POTENTIAL).

FT MOD\_RES 257 257 PHOSPHORYLATION (BY CK2)

FT MOD\_RES 259 259 PHOSPHORYLATION (BY CK2)

FT SEQUENCE 460 AA; 49192 MW; 8A1686C82F5B02E CRC64;

Query Match 28.7%; Score 51; DB 1; Length 460;

Best Local Similarity 38.5%; Pred. No. 24;

Matches 15; Conservative 1; Mismatches 13; Indels 10; Gaps 1;

Qy 1 NPFRTPVQPAFST-----TVPFQSOPVCFPPRGR 29

Db 200 DIAFPVPAFSSPAVAGAAAPASAAYAPPRLGR 238

RESULT 12

TPSP2\_HUMAN ID TSP2\_HUMAN STANDARD PRT; 1172 AA.

AC P35442; DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE THROMBOSPDIN 2 PRECURSOR.

GN TSP2 OR TSP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.

NCBI\_TAXID 9606;

OX [1]

RP SEQUENCE FROM N.A. MEDLINE=94010892; PubMed=8406456;

RA Labeli T.L.; Byers P.H.;

RT "Sequence and characterization of the complete human thrombospondin 2 cDNA: Potential regulatory role for the 3' untranslated region." / Genomics 17:225-229 (1993).

RN [2]

RP SEQUENCE OF 560-1172 FROM N.A.

RC TISSUE-Fibroblast; MEDLINE=92217964; PubMed=1559684;

RA Labeli T.L.; McCookey Milwicz D.J.; Distech C.M.; Byers P.H.;

RT "Thrombospondin II: Partial cDNA sequence, chromosome location, and expression of a second member of the thrombospondin gene family in humans." / Genomics 12:421-429 (1992).

RL -I- FUNCTION: ADHESIVE GLICOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.

CC -I- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS.

CC -I- SIMILARITY: BELONGS TO THE THROMBOSPDIN FAMILY.

CC -I- SIMILARITY: CONTAINS 1 WFPC DOMAIN.

CC -I- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

CC -I- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC -I- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: L12350; AAA03703; 1; .

DR EMBL; M81399; ; A41713.

DR PIR; A42173; A41713.

DR HSSP; P00740; LIXA.

DR MM; 18061; .

DR InterPro; IPR000561; .

DR InterPro; IPR000884; .

DR InterPro; IPR001007; .

DR Pfam; PF00008; EGF; 2; .

DR Pfam; PF00090; tsp\_1; 3.

DR Pfam; PF0003; vwc; 1.

DR PROSITE; PS00022; EGF; 1; FALSE\_NEG.

DR PROSITE; PS01186; EGF; 2; 1.

DR PROSITE; PS00092; TSP; 3.

DR PROSITE; PS01208; VWC; 1.

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat; EGF-like domain; Signal.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 1172 THROMBOSPDIN 2.

FT DOMAIN 19 232 HEPARIN BINDING (POTENTIAL).

FT DOMAIN 318 375 VWF.

FT DOMAIN 381 432 TSP TYPE-1 1.

FT DOMAIN 437 493 TSP TYPE-1 2.

FT DOMAIN 494 548 TSP TYPE-1 3.

FT DOMAIN 549 589 EGF-LIKE 1.

FT DOMAIN 590 647 EGF-LIKE 3.

FT DOMAIN 648 692 EGF-LIKE.

FT DOMAIN 725 760 TSP TYPE-3 1.

FT DOMAIN 761 783 TSP TYPE-3 2.

FT DOMAIN 784 819 TSP TYPE-3 3.

FT DOMAIN 820 842 TSP TYPE-3 4.

FT DOMAIN 843 880 TSP TYPE-3 5.

FT DOMAIN 881 916 TSP TYPE-3 6.

FT DOMAIN 917 952 TSP TYPE-3 7.

FT DOMAIN 953 1172 C-TERMINAL.

FT SITE 928 940 CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 266 266 INTERCHAIN (PROBABLE).

FT DISULFID 270 270 INTERCHAIN (PROBABLE).

FT DISULFID 553 564 BY SIMILARITY.

FT DISULFID 558 574 BY SIMILARITY.

FT DISULFID 577 588 BY SIMILARITY.

FT DISULFID 594 610 BY SIMILARITY.

FT DISULFID 601 619 BY SIMILARITY.

FT DISULFID 622 646 BY SIMILARITY.

FT DISULFID 652 665 BY SIMILARITY.

FT DISULFID 659 678 BY SIMILARITY.

FT DISULFID 680 691 BY SIMILARITY.

FT CARBOHYD 151 151 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 316 316 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 457 457 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 584 584 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 710 710 N-LINKED (GLCNAC . . .) (POTENTIAL).

FT CARBOHYD 1069 1069 N-LINKED (GLCNAC . . .) (POTENTIAL).

SQ SEQUENCE 1172 AA; 129955 MW;

Query Match 28.7%; Score 51; DB 1; Length 1172;

Best Local Similarity 4.8%; Pred. No. 60;

Matches 13; Conservative 2; Mismatches 8; Indels 6; Gaps 2;

Qy 2 PRTVPVQAFSTVFSSPVCFPPRPRRR 30

Db 608 PRCVNTQPGFHCLP----C-PPYRGNQ 630

RESULT 13

L112\_CAEEL STANDARD; PRT; 1429 AA.

ID L112\_CAEEL

AC P14585;

DR 01-JAN-1990 (Rel. 13, Created)

DR 01-JAN-1990 (Rel. 13, Last sequence update)

DR 01-OCT-2000 (Rel. 40, Last annotation update)

DR LIN-12 PROTEIN PRECURSOR.

GN LIN-12 OR R107.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Peloderae; Caenorhabditis.

NCB\_TaxID=6239;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=88334747; PubMed=3419531;

RA Yochim J., Weston K., Greenwald I., "The *Caenorhabditis elegans* lin-12 gene encodes a transmembrane protein with overall similarity to *Drosophila Notch*.", *Nature* 335:547-550(1988).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Croxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laissier N., Latrellie P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Riffkin L., Roopa A., Saunders D., Showken R., Sims M., Smaardon N., Smith A., Smith M., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaughan K., Watson R., Watson A., Wainstock L., Wilkinson-Sproat J., Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans";

RL Nature 368:32-38(1994).

-!- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT REQUIRES CELL-CELL INTERACTIONS. IT IS KNOWN THAT LIN-12 ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.

CC -!- SIMILARITY: HIGH, TO C-ELEGANS GIP-1.

CC -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 3 LIN-NOTCH REPEATS.

CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; MI:2069; AAA:01911; -;

DR EMBL; Z14092; CAA:84741; -;

DR PIR; S06434; S06434;

DR HSSP; P00740; 1IXA;

DR WormPep; R107.8; CE00274.

DR InterPro; IPR000152; -;

DR InterPro; IPR00561; -;

DR InterPro; IPR00880; -;

DR InterPro; IPR001881; -;

DR InterPro; IPR002110; -;

DR Pfam; PF00008; EGF; 13.

DR Pfam; PF00023; ank; 4.

DR Pfam; PF00066; notch; 3.

DR PROSITE; PS50088; ANK\_REPEAT; 3.

DR PROSITE; PS50297; ANK\_REPEAT; 1.

DR PROSITE; PS00010; ASX\_HYDROXYL; 3.

DR PROSITE; PS00022; EGF; 12.

DR PROSITE; PS01186; EGF; 2; 11.

DR PROSITE; PS01187; EGF; 2A; 2.

KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal.

FT SIGNAL 1 15

FT CHAIN 16 1429

FT DOMAIN 16 908

FT TRANSMEM 909 931

FT DOMAIN 932 1429

FT DOMAIN 20 61

FT DOMAIN 114 150

FT DOMAIN 152 190

FT DOMAIN 201 246

FT DOMAIN 250 285

FT DOMAIN 287 323

FT DOMAIN 323 363

FT DOMAIN 365 402

FT DOMAIN 404 441

FT DOMAIN 449 492

FT DOMAIN 503 541

FT DOMAIN 543 579

FT DOMAIN 582 619

FT REPEAT 635 669

FT REPEAT 670 710

FT REPEAT 711 750

FT REPEAT 1093 1126

FT REPEAT 1126 1158

FT REPEAT 1162 1194

FT REPEAT 1206 1236

FT DISULFID 224 35

FT DISULFID 29 49

FT DISULFID 51 60

FT DISULFID 118 129

FT DISULFID 123 138

FT DISULFID 1240 1269

FT DISULFID 140 149

FT DISULFID 156 169

FT DISULFID 163 178

FT DISULFID 180 189

FT DISULFID 205 227

FT DISULFID 221 234

FT DISULFID 236 245

FT DISULFID 254 264

FT DISULFID 259 273

FT DISULFID 275 284

FT DISULFID 291 302

FT DISULFID 296 311

FT DISULFID 313 322

FT DISULFID 327 339

FT DISULFID 334 351

FT DISULFID 353 362

FT DISULFID 369 381

FT DISULFID 375 390

FT DISULFID 392 401

FT DISULFID 408 419

FT DISULFID 413 429

FT DISULFID 431 440

FT DISULFID 507 518

FT DISULFID 512 529

FT DISULFID 531 540

FT DISULFID 547 558

FT DISULFID 552 567

FT DISULFID 569 578

FT DISULFID 586 597

FT DISULFID 591 607

FT DISULFID 609 618

FT CARBOHYD 41 41

FT CARBOHYD 165 165

FT CARBOHYD 194 194

FT CARBOHYD 378 378

-----

N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 515 515 N-LINKED (GLCNAC . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC . .) (POTENTIAL).  
 FT CARBOHYD 751 751 N-LINKED (GLCNAC . .) (POTENTIAL).  
 FT CARBOHYD 754 754 N-LINKED (GLCNAC . .) (POTENTIAL).  
 FT CARBOHYD 900 900 N-LINKED (GLCNAC . .) (POTENTIAL).  
 SQ SEQUENCE 1429 AA; 157115 MW; 255DD7A62C025DB CRC64; [1]

Query Match 28.7%; Score 51; DB 1; Length 1429;  
 Best Local Similarity 54.5%; Pred. No. 72;  
 Matches 12; Conservative 0; Mismatches 6; Indels 4; Gaps 1;

Qy 12 STYPPFSQPVCFF---PPRPRGR 29  
 Db 196 STYEFKQPVCFELISADHPDGR 217

RESULT 14

ID HSB7\_RAT STANDARD PRT; 90 AA.  
 AC Q9QK5; [1]

DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE HEAT-SHOCK PROTEIN, BETA-7 (CARDIOVASCULAR HEAT SHOCK PROTEIN) (CVHSP) (FRAGMENT).  
 GN HSPB7 OR CVHSP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Heart:  
 RX MEDLINE=20062883; PubMed=10593960;  
 RA Krief S., Fairvre J.-F., Robert P., Le Douarin B., Brument-Larignon N., Lefrere I., Bouzyk M.M., Anderson K.M., Greller L.D., Tobin F.L., Souchat M., Bril A.; [1]

RA "Identification and characterization of cvhsp. A novel human small stress protein selectively expressed in cardiovascular and insulin sensitive tissues."  
 RT J. Biol. Chem. 274:3692-36600(1999).

CC -1- SUBUNIT: INTERACTS WITH C-TERMINAL DOMAIN OF ACTIN-BINDING PROTEIN 280 (BY SIMILARITY): FOUND IN BOTH CARDIAC AND SKELETAL MUSCLE.  
 -1- TISSUE SPECIFICITY: FOUND IN BOTH CARDIAC AND SKELETAL MUSCLE.  
 -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ243193; CAB63268.1; [1]

DR PROSITE; PS01031; HSP20; PARTIAL.  
 KW Heat shock; Chaperone.

FT DOMAIN 1 1 8 POLY-SER.  
 FT NON\_TER 90 90 1 8 POLY-SER.

SQ SEQUENCE 90 AA; 9804 MW; 9D122DD6443ED1FE CRC64; [1]

Query Match 28.4%; Score 50.5; DB 1; Length 90;  
 Best Local Similarity 33.3%; Pred. No. 5.7;  
 Matches 11; Conservative 6; Mismatches 7; Indels 9; Gaps 1;

Query Match 28.4%; Score 50.5; DB 1; Length 489;  
 Best Local Similarity 44.4%; Pred. No. 29;  
 Matches 12; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

QY 1 NPR--TVPVQPAFSTVPPFSQPVCFPP 24  
| | | | | | | | | | | | | | | | | | | | | |  
Db 462 NPEDIDTTFVQPGULLSVPPPFEFCFIP 488

Search completed: August 1, 2001, 09:42:18  
Job time: 683 sec



LOCUS AW886539 780 bp mRNA EST 23-MAY-2000 GSS 22-NOV-1998  
 DEFINITION RC1-OT0083-2203000-021-c02 OT0083 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AW886539  
 VERSION AW886539.1 GI:8048551  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Homo; Euteleostomi;  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homo; Homo.  
 REFERENCE 1 (bases 1 to 780)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho A.F., Matsuoka,A., Baia,G.S., Simpson,D.H., Brunsstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 2020663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rue Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=at2-Rc1-0T0083-2203-021-c02&t3=200-03-22&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 26  
 High quality sequence stop: 737.  
 Location/Qualifiers 1..780  
 /organism="Homo sapiens"  
 /db\_xref="taxon:3606"  
 /clone.lib="OT0083"  
 /dev\_stage="Adult"  
 /note="Organ: ovary; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 159 a 265 c 185 g 170 t 1 others  
 ORIGIN  
 alignment\_scores: Quality: 171.00 Length: 32  
 Ratio: 5.516 Gaps: 0  
 Percent Similarity: 96.875 Percent Identity: 96.875  
 alignment\_block: US-09-485-951-1 x AW886539  
 Align seg 1/1 to: AW886539 from: 1 to: 780  
 1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17  
 288 AACCCCGGACAGGCCGTGTCAGCCTGCTTCACGGTGCCTCTC 337  
 17 rGlnProValCysPheProProArgProArgGlyArgArgGlnLys 32  
 338 CCAGCTGCTGTCACCCAGGCCAGGCGCAGACTAAA 383  
 seq\_name: gb\_est4:AW373817  
 seq\_documentation\_block:  
 LOCUS AW373817 665 bp mRNA EST 04-FEB-2000  
 DEFINITION QV3-BU0537-221299-048-c06 BT0537 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION AW373817  
 VERSION AW373817.1 GI:5878471  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





Site_2: NotI; Cloned unidirectionally; Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"	BASE COUNT ORIGIN	219 a 251 c 237 g 201 t 1 others
alignment_scores: Quality: 106.00 Ratio: 4.609 Percent Similarity: 85.185	Length: Gaps: Identity: 74.074	27 0
alignment_block: US-09-485-951-1 x BF181157 ..	from: 1 to: 909	
Align seg 1/1 to: BF181157		
6 PROVALGInPROAlapheserThrValProPheSerGlnProValCysSer 22      :     :     :     :     :     :     :     :     :    488 CCTGTCCAGGCCGTGTCtCTCCACAGTCAGTCACTCAGCCAGTCAGT 537		
22 ePROProArgPzOArgGlyArgArgGlnLys 32      :     :     :     :     :     :     :     :     : 538 CCCACGGACCCCTAAGGGGGAAACAGAAA 568		
seq_name: gb_est89:BF533977		
seq_documentation_block:		
LOCUS BF333977 1165 bp mRNA DEFINITION 602075139F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4212273 5', mRNA sequence.	EST EST EST EST	11-DEC-2000
BF333977 BF333977.1 GI:11621340 EST. house mouse.		
ORGANISM Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 1165)		
COMMENT NIH-MGC http://mgc.nci.nih.gov/.		
REFERENCE AUTHORS TITLE JOURNAL COMMENT		
DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov Plate: LLM9/82 row: m column: 10 High quality sequence stop: 735. Location/Qualifiers		
FEATURES source		
/organism="Mus musculus" /strain="FVB/N" /db_xref="IMAGE:4212273" /clone="NCI_CGAP_Li9" /clone_id="NCI_CGAP_Li9" /label="DH10B (T1 phage-resistant)" /note="Organ: liver; Vector: pCMV-SPRT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library." BASE COUNT ORIGIN	Length: Gaps: Ratio: 4.609	27 0

```

Percent Similarity: 85.185 Percent Identity: 74.074

alignment_block:
US-09-445-951-1 x BF533977 ..

Align seg 1/1 to: BF533977 from: 1 to: 1165

 6 ProValGlnProAlaPheSerThrValProPheSerGlnProValC
 7 |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
 8 491 CCTGAGCTGCTGCTCTCCACAGTGCAGTTCTCAGGCCAGTCC
 9 541 CCCAGGACCCCTAAGGGGCAACAGAAA 571

 22 eProProArgProArgGlyArgArgGlnLys 32
 23 |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
 24 seq_name: gb_est82-BF046316

seq_documentation_block:
 25 LOCUS BF046316 323 bp mRNA EST
 26 DEFINITION BP250002B10C9 Soares normalized bovine plac
 27 clone BP250002B10C9 5', mRNA sequence.
 28 ACCESSION BF046316
 29 VERSION BF046316.1 GI:10763371
 30 KEYWORDS EST.
 31 SOURCE
 32 ORGANISM Bos taurus Metazoa; Chordata; Craniata; Ver-
 33 Mammalia; Eutheria; Cetartiodactyla; Ruminia
 34 Bovidae; Bovinae; Bos.
 35 REFERENCE 1 (bases 1 to 323)
 36 AUTHORS Lewin,H.A., Soares,M.B., Rebeiz,M., Pardina
 37 ,J.H.
 38 TITLE Bovine ESTs
 39 JOURNAL Unpublished (2000)
 40 COMMENT Contact: Lewin, H. A.
 41 W. M. Keck Center for Comparative and Funct.
 42 University of Illinois at Urbana-Champaign
 43 340 Edward R. Madigan Laboratory, 1201 W. G.
 44 61801, USA
 45 Tel: 217 333 5998
 46 Fax: 217 44 5617
 47 Email: h-lewin@uiuc.edu
 48 FORWARD: TAATCGACTCTATAGG
 49 BACKWARD: ATTAAACCTCTACAAAG
 50 Insert Length: 323 Std Error: 0.00
 51 Plate: BP250002B10 Row: C column: 9
 52 Seq Primer: AGCGGATAAACTAACATTCTACAGAGGA
 53 High quality sequence stop: 323.
 54 Location/Qualifiers
 55 1. .323
 56 /organism="Bos taurus"
 57 /db_xref="taxon:9513"
 58 /clone="BP250002B10C9"
 59 /clone_id="Soares normalized bovi
 60 /sex="female"
 61 /lab_host="DH10B"
 62 /note="Organ: placenta; Vector: PT
 63 Site_2: NotI; The cDNA library was
 64 Soares laboratory and it was cons
 65 as described by Bonaldo, M.F., Le
 66 M.B. (1996), Genome Research 6 (9): 59 a 132 c 66 g 63 t 59 g
 67 BASE COUNT
 68 ORIGIN
 69 alignment_scores:

```





```

<clone="IMAGE:4071744"
  /clone_lib="NIH_MGC_76"
  /lab_host="DH10B (T1 phage-resistant)">
<note>"Organ: kidney; Vector: pDNR LIB (Clontech); Site_1: SfiI (ggcccttcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTGGCC-3', and 3' adaptor sequence: 5'-ATCTAGGGCCATTATGCC-3', and 3' adaptor sequence: 5'-ATCTAGGGCCAGGGCACATG-dT(30)BN-3', (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."</note>
  BASE COUNT 271 a 494 c 186 g 130 t
  ORIGIN
  alignment_scores:
    Quality: 70.50 Length: 27
    Ratio: 3.188 Gaps: 1
    Percent Similarity: 51.064 Identity: 48.148
  alignment_block:
    US-09-485-951-1 x BG402468
    Align seg 1/1 to: BG402468 from: 1 to: 1081
    2 ProArgThrValProValGlnProAla...ProSerThrValProPheSer 17
    ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||| 849 CCAAAACATACATCAAACACACCCACACCGAGAACGGCACCAA 898
    531 CCGGGGACGACCT.....TCGGTTATTCCCTTTCCCA 565
    17 rGlnProValCysPheProProArgProArg 27
    :::::||| :::::||| 899 GGAGCCAAACGGCGCACGCCGGCCCCAGA 929
    18 n.....P 19
    566 GAGCAAACGGCCGAAGTGAATGGGGGAGAGCACATCCAGGC 615
    19 rovalCysPheProProArgProArgGlyArgArgGlnLys 32
    |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||| 616 CTCGGGATTCCCTAACCTAGGGCGAAGCAAAG 656
  seq_name: gb_est100:BG402468
  seq_documentation_block:
    LOCUS BG402468 1081 bp mRNA EST 12-MAR-2001
    DEFINITION mRNA sequence.
    ACCESSION BG402468
    KEYWORDS EST.
    SOURCE human.
    ORGANISM Homo sapiens
    Bacteria; Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE NIH-MGC http://mgc.ncbi.nih.gov/.
    AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
    TITLE Unpublished (1999)
    JOURNAL Contact: Robert Strausberg, Ph.D.
    COMMENT Email: cgs@psr@mail.nih.gov
    TISSUE Procurement: CLONTECH Laboratories, Inc.
    CDNA Library Preparation: CLONTECH Laboratories, Inc.
    CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium at:
    http://image.llnl.gov
    Plate: LUCM135 row: m column: 18
    FEATURES High quality sequence stop: 191.
    source Location/qualifiers
    1..1081
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4594361"
    /clone_lib="NIH_MGC_75"

```